

ALMA MATER STUDIORUM
UNIVERSITA' DI BOLOGNA

Scuola di Scienze

Corso di Laurea Magistrale in Matematica

Valutazione di derivati
Americani in modelli
multi-dimensionali

Tesi di Laurea in Equazioni Differenziali Stocastiche

Relatore:

Chiar.mo Prof.

ANDREA PASCUCCI

Presentata da:

ANDREA ANNECCHINI

Anno Accademico 2018/2019

Indice

Introduzione	3
1 Pricing derivati americani	4
1.1 Formulazione del problema	4
1.2 Dinamica dei sottostanti	6
1.2.1 Sottostanti correlati	9
1.3 Frontiera di esercizio ottimo	11
1.4 LSMC	12
1.5 Programmazione dinamica	16
1.6 Algoritmo basato sulla regressione	19
1.6.1 LSM	20
1.7 Bias LSM	20
1.7.1 LOOLSM	22
1.7.2 Convergenza bias	25
1.8 Convergenza coefficienti di regressione	26
1.8.1 Singolo Periodo	26
1.8.2 Multiperiodo	33
1.9 Opzioni Americane	36
1.10 Metodo Monte Carlo	37
1.10.1 Cenni storici	38
1.10.2 Principi matematici	38
2 Esempi Numerici	40
2.1 Call Americana con un sottostante	40
2.2 Esempi numerici con più sottostanti	42
2.2.1 Massimo di due sottostanti indipendenti	42
2.2.2 Massimo di due sottostanti correlati	46
2.2.3 Opzione Call su Basket di quattro sottostanti	50
2.3 Convergenza coefficienti di regressione	52
Appendice	54
Bibliografia	57

Introduzione

I derivati con opportunità di esercizio anticipato solo quelli di maggiore interesse in ambito finanziario. Il pricing di questo tipo di opzioni è un problema molto difficoltoso in quanto, anche nei casi più semplici di opzioni su un singolo sottostante, non esistono soluzioni in forma chiusa.

La ricerca dunque ha portato a sviluppare diversi metodi numerici per il pricing che a grandi linee si dividono in due categorie: quelli basati sulla simulazione e quelli basati sui reticoli.

In questo secondo caso, il pricing viene effettuato su un reticolo denso nello spazio degli stati, valutando l'opzione in ogni suo punto. In questa tipologia di metodi rientrano per esempio quello degli alberi binomiali, lo schema delle differenze finite e le tutte le loro generalizzazioni per problemi multidimensionali. Questi metodi funzionano bene per problemi che hanno una bassa dimensionalità, questo perchè all'aumentare del numero di variabili di stato la grandezza del reticolo cresce esponenzialmente e questo crea problemi dal punto di vista computazionale.

L'approccio basato sulla simulazione invece richiede il calcolo dell'attesa dei valori dell'opzione ognuno fatto in base alla realizzazione futura della variabile di stato rispetto alla misura martingala (o neutrale al rischio). In questa categoria di metodi rientrano il metodo delle foreste casuali e della mesh stocastica. In questo elaborato verrà posta l'attenzione sul metodo LSM, anch'esso basato sulla simulazione, che approssima il valore di continuazione dell'opzione utilizzando la regressione.

Il testo è così strutturato: nel capitolo 1 viene formulato il problema del pricing di un derivato americano e viene mostrato come si simulano le traiettorie dei sottostanti. Sempre nello stesso capitolo viene presentato il metodo LSM e una sua generalizzazione che permette di eliminare gli high bias particolarmente problematici nelle Callable Note Markets. Infine viene trattata la convergenza dei coefficienti di regressione e della stima del valore di continuazione in relazione al numero di simulazioni e al numero di basi utilizzate nella regressione. Il capitolo 2 invece presenta numerosi esperimenti numerici fatti utilizzando entrambi gli algoritmi trattati. Nei test numerici sono stati consideranti diversi tipi di opzioni e vengono fatte considerazioni sulla convergenza, sull'errore e sul costo computazionale degli algoritmi.

Capitolo 1

Pricing derivati americani

A differenza di un'opzione europea che può essere esercitata solo alla data di scadenza fissata, un'opzione americana può essere esercitata in qualsiasi momento fino alla sua scadenza. Essendo un'opzione con possibilità di esercizio anticipato, il calcolo del valore iniziale dell'opzione richiede la risoluzione del problema di quando esercitarla ottimalmente.

Vediamo come viene formulato questo problema e poi come può essere risolto.

1.1 Formulazione del problema

Se si vuole prestare la propria attenzione su una classe molto generale di problemi di pricing di derivati americani che hanno scadenza al tempo T , si può prendere in considerazione il processo $\{U_t, 0 \leq t \leq T\}$, che rappresenta il payoff ottenuto dall'esercizio dell'opzione al tempo t scontato al tempo 0, e l'insieme dei tempi di arresto ammissibili \mathcal{T} con valori in $[0, T]$.

Il problema di arresto ottimale consiste nel trovare:

$$\sup_{\tau \in \mathcal{T}} E[U_\tau]$$

cioè si è interessati al massimo valore atteso del payoff scontato ottenuto stoppando il processo con un tempo di arresto τ ammissibile.

Con un piccola perdita di generalità si può considerare una formulazione più semplice che prende in considerazione il valore del sottostante $\{S_t, 0 \leq t \leq T\}$ come un processo di Markov a valori in \mathbb{R}^d . Questo vuol dire che fino al tempo t conosco tutte le informazioni relative al passato per quanto riguarda le variabili finanziarie che vengono prese in esame e dunque anche il valore dell'opzione nei tempi precedenti.

Indicando con $h(S_t)$ il payoff dell'opzione relativo all'esercizio al tempo t , dove h è una funzione non negativa, e se si considera l'esistenza di un titolo privo di rischio con capitalizzazione continua con un tasso di interesse

costante r allora il problema si riconduce a calcolare:

$$\sup_{\tau \in \mathcal{T}} [e^{-r\tau} h(S_\tau)] \quad (1.1)$$

Si vuole risolvere questo problema attraverso la simulazione. I valori dei sottostanti non possono essere simulati nell'intervallo continuo $[0, T]$, ma abbiamo bisogno di ricorrere ad una discretizzazione dell'intervallo. Quindi in realtà si considera il pricing di un'opzione americana che può essere esercitata solo in un numero finito m di istanti temporali. Questo tipo di opzioni vengono di solito chiamate bermudiane perchè sono a metà tra un'opzione europea e una americana. L'idea è quella di prezzare un'opzione bermudiana e considerando un m abbastanza grande dare un'approssimazione del valore dell'opzione americana continua.

Si consideri dunque un'opzione che può essere esercitata solo in m istanti temporali:

$$0 = t_0 < t_1 < \dots < t_{m-1} < t_m = T$$

È ipotesi usuale supporre che l'opzione non può essere esercitata al tempo 0. Di seguito vengono riportate alcune notazioni che sono utilizzate all'interno di tutto il testo :

- b indica il numero di simulazioni
- S_{ij} indica il valore del sottostante al tempo t_i relativo alla j -esima simulazione
- $S_i = (S_{i1}, S_{i2}, \dots, S_{ib})$ indica il vettore dei valori del sottostante al tempo t_i
- $h_i(S_{ij})$ indica il payoff che paga l'opzione se viene esercitata al tempo t_i relativamente alla simulazione j .

$$h_i : \mathbb{R}^d \rightarrow [0, +\infty) \quad i = 1, 2, \dots, m$$

- $h_i(S_i) = (h_i(S_{i1}), h_i(S_{i2}), \dots, h_i(S_{ib}))$ indica il vettore dei payoff al tempo t_i
- \mathcal{T}_i indica l'insieme dei tempi di arresto con valori in $\{t_i, t_{i+1}, \dots, t_m\}$

Prima di presentare il funzionamento degli algoritmi, viene mostrato come sono modellizzati i sottostanti e in che modo si simulano le traiettorie degli stessi.

1.2 Dinamica dei sottostanti

Il nostro modello prevede che il valore del sottostante S soddisfi la seguente equazione differenziale stocastica:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (1.2)$$

Dove μ e $\sigma > 0$ sono due costanti e W_t indica un moto Browniano standard. La soluzione di questa equazione differenziale stocastica è detta Moto Browniano geometrico ed è data esplicitamente da:

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

Possiamo verificare che questa è effettivamente la soluzione ricorrendo alla formula di Itô. Sia

$$F(t, x) = e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma x} \in C^{2,1}(\mathbb{R}^2)$$

Calcoliamo le derivate di F che ci servono

$$F_x(t, x) = \sigma F(t, x), \quad F_{xx}(t, x) = \sigma^2 F(t, x)$$

$$F_t(t, x) = \left(\mu - \frac{1}{2}\sigma^2 \right) F(t, x)$$

Dalla formula di Itô abbiamo che

$$\begin{aligned} dS_t &= dF(t, W_t) = \left(F_t(t, W_t) + \frac{1}{2}F_{xx}(t, W_t) \right) dt + F_x(t, W_t)dW_t = \\ &= \left(\mu - \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 \right) F(t, W_t)dt - \sigma F(t, W_t)dW_t \\ &= \mu F(t, W_t)dt + \sigma F(t, W_t)dW_t = \mu S_t dt + \sigma S_t dW_t \end{aligned}$$

I due parametri μ e σ vengono chiamati rispettivamente drift e volatilità. Il primo indica il valore medio del rendimento annuale del titolo. Il secondo parametro invece ha una grande importanza perchè indica di quanto il titolo oscilla, infatti esso non è altro che la deviazione standard del rendimento annuale. Si può notare che se $\sigma = 0$ non è più presente la parte aleatoria e quindi in questo caso il sottostante presenta una dinamica deterministica:

$$S_t = S_0 e^{\mu t}$$

Prima di dedicarsi al pricing di un derivato risulta importante stimare questi due parametri che servono per modellizzare la dinamica del sottostante. Questa fase viene chiamata calibrazione. La volatilità ha un ruolo estremamente importante e non risulta facile nei mercati reali darne una stima. Di solito questa si basa sullo studio della storia del sottostante e sull'utilizzo

delle serie storiche. Una volta che questi parametri sono stati stimati il valore del sottostante viene modellizzato con un moto Browniano geometrico e si procede con la simulazione di traiettorie e all'applicazione degli algoritmi di pricing.

Inizialmente Bachelier utilizzò il moto Browniano standard per modellizzare i prezzi dei titoli e solo successivamente Samuelson considerò il moto Browniano geometrico più appropriato. Il perchè di questa scelta risiede nel fatto che il moto Browniano geometrico assume sempre valori positivi, a differenza del moto Browniano che con probabilità positiva può assumere valori negativi. Questo risulta importante perchè non è molto sensato dare un valore negativo al sottostante. Il secondo deriva dal fatto che gli incrementi relativi cioè

$$\frac{S_1 - S_0}{S_0}, \frac{S_2 - S_1}{S_1}, \dots, \frac{S_{t_n} - S_{t_{n-1}}}{S_{t_{n-1}}}$$

nel moto Browniano geometrico risultano dipendenti.

Questa prima dinamica che è stata considerata è molto semplice. Si possono prendere in esame anche casi più complicati. Per esempio si può considerare un drift non costante ma come funzione deterministica del tempo $\mu(t)$ e in questo caso il sottostante soddisfa la seguente equazione differenziale stocastica

$$dS_t = \mu(t)S_t dt + \sigma S_t dW_t$$

e la dinamica è data da

$$S_t = S_0 e^{(\int_0^t \mu(s) ds - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

Oppure generalizzando ancora si può considerare il drift, ma soprattutto la volatilità come dei processi stocastici. In questo caso si dice che la volatilità è endogena se è descritta da un processo che dipende dagli stessi fattori di rischio del sottostante. Se invece la volatilità è legata a fattori di rischio aggiuntivi si dice che è esogena. Generalmente nel primo caso si ha un modello di mercato che è completo mentre nel secondo incompleto. Tra i modelli endogeni hanno particolare importanza quelli chiamati a volatilità locale, cioè dove si assume che σ e μ sono funzioni del tempo e del valore del sottostante. Dunque in questo caso la dinamica del sottostante è soluzione di:

$$dS_t = \mu(t, S_t)S_t dt + \sigma(t, S_t)S_t dW_t$$

Noi ci limiteremo a considerare solo il caso in cui la dinamica è data da (1.2). Prendiamo in esame un modello di mercato con tasso di interesse privo di rischio costante r su un intervallo limitato $[0, T]$. I dividendi nello studio delle opzioni americane hanno un ruolo fondamentale e per questo nei nostri casi il sottostante avrà la seguente dinamica :

$$dS_t = (r - \delta)S_t dt + \sigma S_t dW_t$$

Dove $\delta \geq 0$ è il tasso di rendimento del dividendo che consideriamo costante nel tempo. In pratica la dinamica è quella di un moto geometrico con drift $r - \delta$ e volatilità σ . Per indicare che il valore del sottostante ha questo andamento usiamo la seguente notazione $S \sim MBG(r - \delta, \sigma)$.

Vediamo ora come si simula una traiettoria del sottostante.

Sia $S \sim MBG(\mu, \sigma)$ e consideriamo $t > u$ si può verificare che la dinamica di S_t può essere scritta anche in questo modo:

$$S_t = S_u e^{(\mu - \frac{1}{2}\sigma^2)(t-u) + \sigma(W_t - W_u)}$$

Questo ci suggerisce di simulare la traiettoria di S in modo ricorsivo. La prima cosa da fare è discretizzare l'intervallo $[0, T]$. Quindi consideriamo

$$0 < t_1 < t_2 < \dots < t_{m-1} < t_m = T$$

i punti di discretizzazione dell'intervallo. Abbiamo m intervalli $[t_{i-1}, t_i]$ che per semplicità si suppongono di uguale lunghezza $dt = \frac{T}{m}$. Siccome il valore iniziale S_0 è osservabile possiamo simulare il valore del sottostante nei punti di discretizzazione mediante la seguente formula ricorsiva:

$$S_i = S_{i-1} e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma(W_i - W_{i-1})} \quad i = 1, 2, \dots, m$$

Essendo W un moto Browniano standard abbiamo che i suoi incrementi sono indipendenti e in particolare hanno distribuzione normale:

$$W_i - W_{i-1} \sim \mathcal{N}_{0,dt} \sim \sqrt{dt}Z \quad i = 1, 2, \dots, m$$

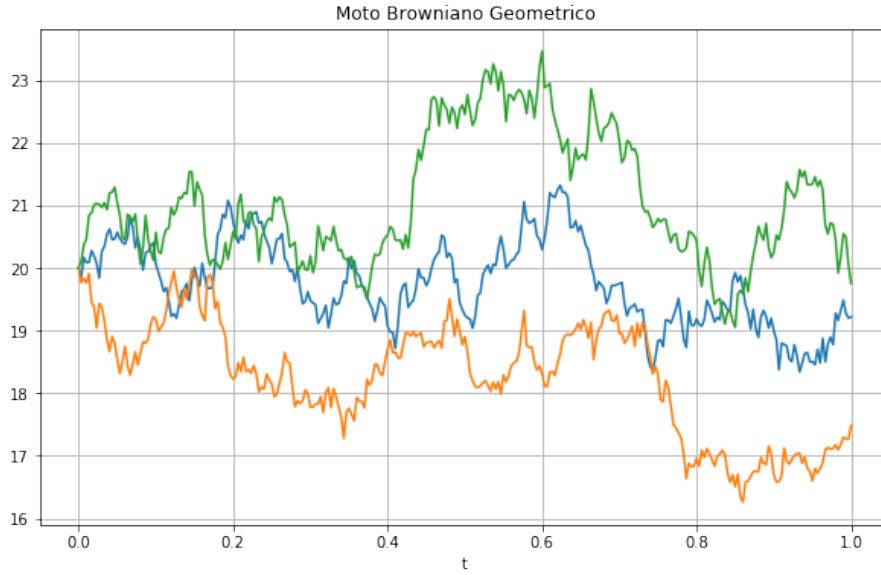
Dove Z indica una normale standard. Perciò la nostra formula ricorsiva diventa

$$S_i = S_{i-1} e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma\sqrt{dt}Z} \quad i = 1, 2, \dots, m$$

Dunque per ogni $i = 1, 2, \dots, m$ realizziamo in modo indipendente una variabile normale standard e ricorsivamente troviamo tutti i valori del sottostante nei punti di discretizzazione.

Di seguito è riportato un esempio con $b = 3$ traiettorie simulate nell'intervallo $[0, 1]$ con un numero $m = 300$ di punti di discretizzazione di un sottostante modellizzato con un moto Browniano Geometrico con i seguenti parametri:

$$\mu = 0.06 \quad \sigma = 0.2 \quad S_0 = 20$$



L'esempio riguarda un solo sottostante ma in generale le opzioni possono dipendere da un numero n di sottostanti. Inoltre essi potrebbero essere legati ad uno stesso evento aleatorio e quindi essere dipendenti tra di loro. Nel caso in cui siano indipendenti basta considerare ogni sottostante modellizzato con un moto Browniano Geometrico, dove i moti Browniani standard di ogni sottostante sono tutti indipendenti tra di loro. Di seguito vediamo la trattazione del caso in cui invece dovessero esseri dipendenti tra loro.

1.2.1 Sottostanti correlati

Data una matrice α di dimensione $n \times d$, definiamo la matrice di correlazione come

$$C = \alpha \alpha^*$$

che è simmetrica e semidefinita positiva. Preso $\eta \in \mathbb{R}^n$ e un moto Browniano d -dimensionale W poniamo

$$B_t = \eta + \alpha W_t$$

cioè

$$dB_t = \alpha dW_t$$

Diciamo che B è un moto Browniano correlato di punto iniziale η e matrice di correlazione C .

Indicando con

- $S_t = (S_t^1, S_t^2, \dots, S_t^n)$ il vettore dei valori dei sottostanti
- $B_t = (B_t^1, B_t^2, \dots, B_t^n)$ le componenti del moto Browniano correlato

- $\mu = (\mu^1, \mu^2, \dots, \mu^n)$ il vettore dei drift dei sottostanti
- $\sigma = (\sigma^1, \sigma^2, \dots, \sigma^n)$ il vettore delle volatilità dei sottostanti

possiamo scrivere in forma compatta la nostra modellizzazione:

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

e in forma esplicita:

$$dS_t^1 = \mu^1 S_t^1 dt + \sigma^1 S_t^1 dB_t^1$$

$$dS_t^2 = \mu^2 S_t^2 dt + \sigma^2 S_t^2 dB_t^2$$

\vdots

$$dS_t^n = \mu^n S_t^n dt + \sigma^n S_t^n dB_t^n$$

Prendiamo ora in esame un caso particolare con due sottostanti. Supponiamo che abbiamo stesso drift e stessa volatilità. Essi soddisfano le seguenti equazioni differenziali stocastiche:

$$dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 dB_t^1$$

$$dS_t^2 = \mu S_t^2 dt + \sigma S_t^2 dB_t^2$$

Con

$$B_t = \alpha W_t$$

un moto Browniano correlato con punto iniziale nullo e matrice di correlazione

$$C = \alpha \alpha^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Dove $\rho \in [-1, 1]$. Da un semplice calcolo si può verificare che

$$\alpha = \begin{pmatrix} \rho & \sqrt{1 - \rho^2} \\ 1 & 0 \end{pmatrix}$$

Dunque che il nostro moto Browniano correlato ha componenti:

$$B = \alpha W = \begin{pmatrix} \rho & \sqrt{1 - \rho^2} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} W^1 \\ W^2 \end{pmatrix} = \begin{pmatrix} \rho W^1 + \sqrt{1 - \rho^2} W^2 \\ W^1 \end{pmatrix}$$

Le equazioni possono essere riscritte in questo modo:

$$dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 (\rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2)$$

$$dS_t^2 = \mu S_t^2 dt + \sigma S_t^2 dW_t^1$$

Con la formula di Itô si può vedere che

$$S_t^1 = S_0^1 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma(\rho W_t^1 + \sqrt{1 - \rho^2} W_t^2)}$$

$$S_t^2 = S_0^2 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t^1}$$

Per simulare i due sottostanti, ragionando in modo simile a prima, otteniamo:

$$S_i^1 = S_{i-1}^1 e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma\sqrt{dt}(\rho Z_1 + \sqrt{1-\rho^2}Z_2)} \quad i = 1, 2, \dots, m$$

$$S_i^2 = S_{i-1}^2 e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma\sqrt{dt}Z_1} \quad i = 1, 2, \dots, m$$

Dove Z_1 e Z_2 sono realizzazioni di due variabili normali standard indipendenti tra di loro.

Dopo che è stato presentato come si possono simulare le traiettorie dei sottostanti, si vuole porre attenzione sulla politica di esercizio e sul concetto di frontiera ottimale di esercizio.

1.3 Frontiera di esercizio ottimo

Abbiamo visto nella sezione 1.1 che il problema principale che si presenta quando siamo di fronte ad un'opzione americana è capire quando essa può essere esercitata in modo ottimale. Questo ci impone di trovare una politica decisionale di esercizio. Prima di presentare il concetto di frontiera di esercizio ottimo abbiamo bisogno di queste due definizioni:

Definizione. Il *valore intrinseco* di un'opzione si definisce come il valore che avrebbe se fosse esercitata immediatamente cioè la differenza tra il prezzo di esercizio e il valore della azione.

Definizione. Il *prezzo critico* è il prezzo del sottostante in corrispondenza del quale il valore dell'opzione è pari al suo valore intrinseco.

Consideriamo un'opzione americana esercitabile nell'intervallo $[0, T]$, la frontiera di esercizio ottimo è una funzione $b^* : [0, T] \rightarrow \mathbb{R}$ tale che $b^*(t)$ rappresenta il prezzo critico del sottostante al tempo t .

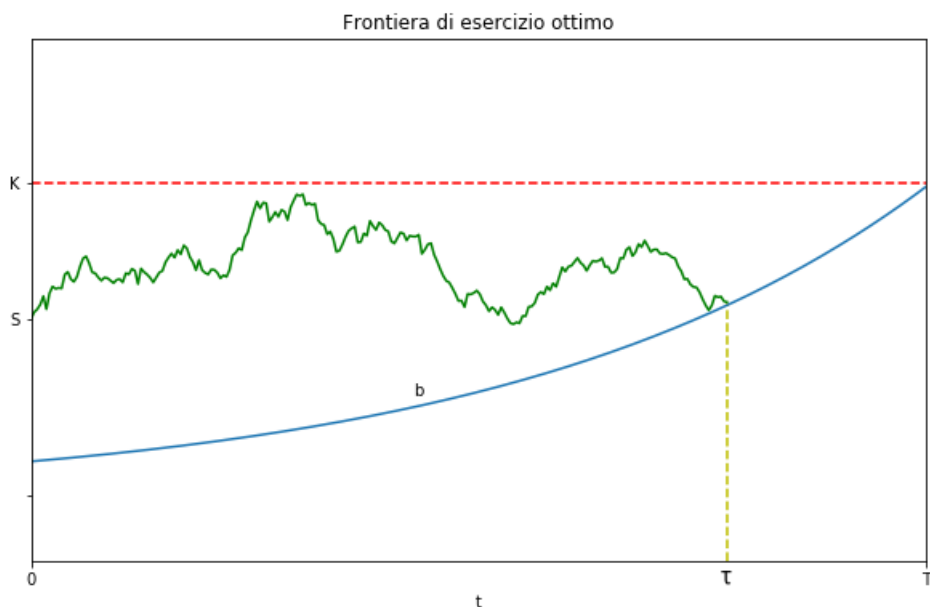
Consideriamo ad esempio un'opzione Put americana, possiamo definire la sua regione di esercizio

$$E = \{(x, \tau) \in \mathbb{R}^+ \times [0, T] : S_\tau = x, x \leq b^*(\tau)\}$$

e la regione di continuazione

$$F = \{(x, \tau) \in \mathbb{R}^+ \times [0, T] : S_\tau = x, x > b^*(\tau)\}$$

Per dare un'idea consideriamo la seguente figura:



Nota la frontiera di esercizio ottimo il tempo di arresto ottimale che realizza il sup (1.1) si può ottenere nella forma:

$$\tau^* = \inf\{t \in [0, T] : S_t \leq b^*(t)\}$$

Di conseguenza conoscendo la frontiera di esercizio ottimo si può dare una stima del valore iniziale dell'opzione. Si simulano delle traiettorie del sottostante e per ognuna si calcola il valore del tempo τ di quando essa tocca la frontiera di esercizio (si pone $\tau = T$ se non la tocca). Tra tutti i tempi calcolati si prende il minimo che possiamo indicare con $\hat{\tau}^*$.

La stima del prezzo iniziale viene calcolata con

$$\hat{V}_0 = e^{-r\hat{\tau}^*} h(S_{\hat{\tau}^*})$$

Il problema dunque si riconduce a dover trovare i valori critici e la frontiera di esercizio ottimo.

Purtroppo anche nei casi più semplici di opzioni, come quelle Put e Call, non esistono forme analitiche di b^* ma solo forme integrali. Quindi per stimare la frontiera bisogna ricorrere a delle integrazioni numeriche oppure ad altri approcci come per esempio la tecnica di randomizzazione utilizzata da Carr (1998).

1.4 LSMC

La tecnica Monte Carlo viene usata in finanza per prezzare le opzioni europee : si simulano un gran numero di traiettorie del valore del sottostante,

per ognuna di queste si calcola il payoff dell'opzione e si approssima la media dei payoff con la media aritmetica e infine si attualizza il valore per ottenere la stima del prezzo iniziale dell'opzione. Questa tecnica non è stata applicata molto per prezzare opzioni americane che come abbiamo visto pongono il problema dell'esercizio ottimo. Carriere nel 1996 mostrò che la stima della scelta di esercizio ottimale dell'opzione era equivalente a stimare una somma di attese condizionate. Tsitsiklis e Van Roy [4] proposero un algoritmo che approssimava le attese condizionate attraverso combinazioni lineari di funzioni base ricorrendo alla regressione ai minimi quadrati per stimare i coefficienti. Anche Longstaff e Schwartz applicarono lo stesso tipo di regressione per stimare i valori di continuazione delle opzioni. Loro diedero a questa tecnica il nome di "Metodo Monte Carlo ai minimi quadrati" (LSMC).

Questo metodo precisamente coinvolge due approssimazioni:

- Sostituire l'attesa condizionata nell'algoritmo di pricing con la sua proiezione ortogonale sullo spazio generato da un insieme finito di funzioni base.
- Usare la simulazione Monte Carlo e la regressione ai minimi quadrati per stimare numericamente la funzione attesa condizionata.

Prima di presentare in dettaglio i principi e il funzionamento del metodo è opportuno fare delle considerazioni iniziali.

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità, data una variabile aleatoria:

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}_d)$$

e una variabile aleatoria in uno spazio misurabile generico:

$$Y : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$$

consideriamo $Z = E[X|Y]$ una versione dell'attesa condizionata di X a Y . Per una delle proprietà dell'attesa condizionata si ha che $Z \in m\sigma(Y)$ (cioè è misurabile rispetto alla sigma algebra generata da Y) e per il teorema di Doob si ha che esiste (e non è detto che sia unica) una funzione Φ misurabile secondo Borel tale che $Z = \Phi(Y)$. Diciamo che Φ è una versione della funzione attesa condizionata di X a Y e utilizziamo la seguente notazione:

$$\Phi = E[X|Y = \cdot]$$

La Φ è una versione della funzione attesa condizionata di X a Y se e solo se $\Phi(Y)$ è una versione dell'attesa condizionata di X a Y . In altri termini possiamo vedere l'attesa condizionata a Y come una funzione oppure come una variabile aleatoria. La scelta di quale delle due rappresentazioni considerare dipende generalmente dal contesto preso in esame.

Vediamo come è caratterizzata l'attesa condizionata nello spazio L^2 .

Teorema 1. Sia X una variabile aleatoria tale che $X \in L^2(\Omega, \mathcal{F}, P)$ e sia \mathcal{G} una sotto- σ -algebra di \mathcal{F} . Data $W \in L^2(\Omega, \mathcal{G}, P)$ si ha che

$$\langle X - E[X|\mathcal{G}], W \rangle_{L^2} = 0$$

Dove $\langle \cdot, \cdot \rangle_{L^2}$ è il prodotto scalare in L^2 .

Dunque $X - E[X|\mathcal{G}]$ è ortogonale al sottospazio $L^2(\Omega, \mathcal{G}, P)$ e quindi l'attesa condizionata $E[X|\mathcal{G}]$ non è altro che la proiezione di X sul sottospazio $L^2(\Omega, \mathcal{G}, P)$. Infatti se $Z = E[X|\mathcal{G}]$ si ha che $Z \in L^2(\Omega, \mathcal{G}, P)$ e per ogni $W \in L^2(\Omega, \mathcal{G}, P)$ vale che

$$E[(X - Z)^2] \leq E[(X - W)^2]$$

In altri termini Z realizza la minima distanza di X da $L^2(\Omega, \mathcal{G}, P)$ e quindi rappresenta geometricamente la sua migliore approssimazione nello spazio $L^2(\Omega, \mathcal{G}, P)$. Nella pratica siamo interessati al caso in cui $\mathcal{G} = \sigma(Y)$ con $Y \in L^2(\Omega, \mathcal{F}, P)$. Per quanto visto prima abbiamo che se $Z = E[X|Y]$ allora $Z = \Phi(Y)$ con $\Phi \in L^2(\mathbb{R}, \mathcal{B}, \mu_Y)$. Siccome Z realizza la minima distanza si ha che

$$E[(X - \Phi(Y))^2] = \min_{f \in L^2(\mathbb{R}, \mathcal{B}, \mu_Y)} E[(X - f(Y))^2]$$

e quindi per trovare Φ bisogna risolvere il problema ai minimi quadrati

$$\Phi = \arg \min_{f \in L^2(\mathbb{R}, \mathcal{B}, \mu_Y)} E[(X - f(Y))^2]$$

Vediamo ora come si può risolvere il problema utilizzando il LSMC. Consideriamo un caso più generale, pensiamo di voler approssimare numericamente l'attesa condizionata

$$\Phi(Y) = E[F(X, Y)|Y]$$

dove $F(X, Y) \in L^2(\Omega, \mathcal{F}, P)$ e siamo a conoscenza della distribuzione congiunta $\mu_{(X, Y)}$. Possiamo notare che se X e Y sono indipendenti, come conseguenza del lemma di freezing, abbiamo che

$$\Phi = E[F(X, \cdot)]$$

e quindi possiamo calcolare questo valore atteso semplicemente con il Metodo Monte Carlo. Per questo è di maggiore interesse il caso in cui X e Y sono dipendenti. La Φ deve risolvere il problema ai minimi quadrati

$$\Phi = \arg \min_{f \in L^2(\mathbb{R}, \mathcal{B}, \mu_Y)} E[(f(Y) - F(X, Y))^2]$$

L'idea del LSMC è di approssimare in dimensione finita il problema. Siccome lo spazio L^2 è uno spazio di Hilbert, possiamo pensare di esprimere la f

come combinazione lineare di funzioni base ortonormali come per esempio i polinomi di Laguerre, di Hermite, di Legendre o di Jacobi. I test numerici però mostrano che anche un' approssimazione in serie di Fourier o semplicemente utilizzando le potenze permette di ottenere lo stesso risultati accurati. Dunque prendiamo le funzioni polinomiali:

$$\beta_k(y) := y^k \quad k = 0, 1, 2, \dots$$

Consideriamo il sottospazio di dimensione $N + 1$ generato dalle funzioni base β_0, \dots, β_N . Indicando con

$$\beta = (\beta_0, \dots, \beta_N)$$

e con

$$\lambda = (\lambda_0, \dots, \lambda_N) \in \mathbb{R}^{N+1}$$

Approssimiamo $f \in L^2(\mathbb{R}, \mathcal{B}, \mu_Y)$ con la sua proiezione sul sottospazio generato da β , cioè:

$$f(y) \approx \langle \beta(y), \lambda \rangle$$

Quindi il nostro problema si riconduce a trovare $\bar{\lambda} \in \mathbb{R}^{N+1}$ che risolve

$$\min_{\lambda \in \mathbb{R}^{N+1}} E[(\langle \beta(Y), \lambda \rangle - F(X, Y))^2]$$

Trovato $\bar{\lambda}$, l'attesa condizionata si approssima con

$$\Phi \approx \langle \beta, \bar{\lambda} \rangle$$

Vediamo ora come possiamo trovare $\bar{\lambda}$. Conoscendo la distribuzione congiunta possiamo simulare b valori delle variabili casuali X e Y . Indichiamo con $x = (x_k)$ e $y = (y_k)$ con $k = 1, \dots, b$ i vettori che hanno come componenti i valori simulati. Il numero di simulazioni b deve essere sufficientemente grande mentre N viene mantenuto piccolo per questioni di convergenza e anche per avere un costo computazionale minore nel risolvere il problema. Approssimiamo il valore atteso utilizzando il Metodo Monte Carlo cioè:

$$E[(\langle \beta(Y), \lambda \rangle - F(X, Y))^2] \approx \frac{\sum_{k=1}^b (\langle \beta(y_k), \lambda \rangle - F(x_k, y_k))^2}{b}$$

Posto $\mathbf{B} = (\beta_i(y_k))$ e $\mathbf{F} = (F(x_k, y_k))$ dove $k = 1, \dots, b$ e $i = 0, \dots, N$ possiamo scrivere in modo più compatto l'approssimazione:

$$E[(\langle \beta(Y), \lambda \rangle - F(X, Y))^2] \approx \frac{|\mathbf{B}\lambda - \mathbf{F}|^2}{b} \quad \lambda \in \mathbb{R}^{N+1}, \quad b \gg 1$$

Dobbiamo minimizzare

$$Q(\lambda) = |\mathbf{B}\lambda - \mathbf{F}|^2$$

Troviamo il minimo imponendo $\nabla Q(\lambda) = 0$.

$$\nabla Q(\lambda) = 2(\mathbf{B}^T \mathbf{B} \lambda - \mathbf{B}^T \mathbf{F})$$

Quindi il minimo $\bar{\lambda}$ risolve

$$\mathbf{B}^T \mathbf{B} \lambda = \mathbf{B}^T \mathbf{F}$$

Vediamo un esempio pratico di come si approssima la funzione attesa condizionata con il LSMC.

Consideriamo le variabili aleatorie X e Y che hanno come distribuzione congiunta una binormale

$$(X, Y) \sim \mathcal{N}_{\mu, C}$$

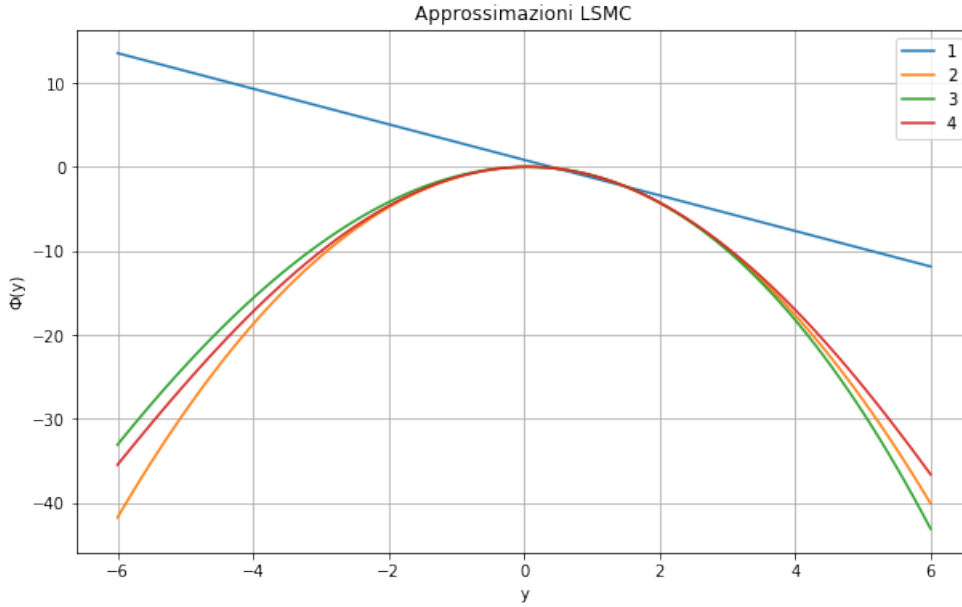
con

$$\mu = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad C = \begin{pmatrix} 0.64 & -0.28 \\ -0.28 & 0.25 \end{pmatrix}$$

Prendiamo $F(x, y) = xy$ e calcoliamo

$$\Phi(Y) = E[F(X, Y)|Y]$$

Di seguito sono riportate le prime quattro approssimazioni LSMC utilizzando la base polinomiale.



1.5 Programmazione dinamica

Vediamo ora come il problema del pricing viene formulato attraverso il LSMC.

Definiamo per $i = 1, 2, \dots, m$

$$V_i(x) = \sup_{\tau \in \mathcal{T}_i} E[h_\tau(S_\tau) | S_i = x]$$

che rappresenta il valore dell'opzione al tempo t_i nello stato x se l'opzione non è stata esercitata prima di t_i . Siccome abbiamo considerato uno sconto deterministico questo può essere inglobato all'interno dell' h in modo da non portarci dietro nei calcoli il termine di sconto. Il valore dell'opzione soddisfa le seguenti equazioni di programmazione dinamica:

$$V_m(x) = h_m(x) \quad (1.3)$$

$$V_i(x) = \max\{h_i(x), E[V_{i+1}(S_{i+1})|S_i = x]\} \quad (1.4)$$

con $i = 1, 2, \dots, m - 1$.

La (1.3) deriva dal fatto che alla scadenza il valore dell'opzione è dato dal payoff in quanto non si può esercitare l'opzione oltre la scadenza. La (1.4) ci dice che il valore dell'opzione al tempo i è il massimo del payoff al tempo i e l'attesa dell'opzione al tempo $i + 1$ (nello stato S_{i+1}) condizionata al valore di S al tempo i . Questo secondo termine non è altro che il valore di continuazione, cioè il valore che ha l'opzione se non viene esercitata.

Indichiamo quindi con

$$C_i(S_i) = E[V_{i+1}(S_{i+1})|S_i]$$

il valore di continuazione. Chiaramente con $C_i(x)$ indichiamo il valore di continuazione al tempo t_i e nello stato $S_i = x$. Anche per il valore di continuazione possiamo scrivere le seguenti equazioni di programmazione dinamica:

$$C_m(x) = 0$$

$$C_i(x) = E[\max\{h_{i+1}(S_{i+1}), C_{i+1}(S_{i+1})\}|S_i = x]$$

con $i = 1, 2, \dots, m - 1$.

La (1.4) può essere quindi riscritta in questo modo:

$$V_i(x) = \max\{h_i(x), C_i(x)\}$$

Siccome i payoff sono funzioni note, per calcolare il valore di V abbiamo bisogno di stimare solo i valori di continuazione e questo può essere fatto ricorrendo ad un'approssimazione finita utilizzando delle funzioni base come abbiamo già visto nella sezione precedente.

Se utilizziamo $N + 1$ funzioni base per ogni $i = 1, 2, \dots, m$ indichiamo con

$$\psi_i = (\psi_{i0}, \psi_{i1}, \dots, \psi_{iN})$$

il vettore delle funzioni base che utilizziamo per stimare il valore di continuazione al tempo t_i , allora la sua approssimazione è data da:

$$C_i(S_i) \approx \sum_{k=0}^N \beta_{ik} \psi_{ik}(S_i)$$

dove β_{ik} sono i coefficienti che dobbiamo trovare per approssimare la funzione.

Indichiamo con

$$\beta_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{iN})$$

il vettore dei coefficienti relativi alla stima del valore di continuazione al tempo t_i e con

$$B_\psi = E[\psi(S_i)\psi(S_i)^T] \quad \gamma = E[\psi(S_i)V_{i+1}(S_{i+1})]$$

Assumiamo che

(A0)

$$\psi_{i0} \equiv 1, \quad E[\psi_i(S_i)] = 0 \quad i = 1, 2, \dots, m$$

e che B_ψ sia finita e non singolare.

Per ogni data di esercizio abbiamo supposto di utilizzare funzioni base differenti, questo però non è necessario e quindi per semplicità possiamo usare le stesse basi per ogni istante temporale di esercizio.

Il vettore β_i è ottenuto utilizzando la regressione ai minimi quadrati dalla proiezione di V_{i+1} sullo span generato da ψ_i , cioè il vettore β_i è dato da:

$$\beta_i = B_\psi^{-1}\gamma$$

Dove B_ψ è una matrice $(N+1) \times (N+1)$ e γ è un vettore di lunghezza $N+1$. Prendiamo in considerazione b traiettorie simulate del valore del sottostante S . La stima dei coefficienti di regressione è data da:

$$\hat{\beta}_i = \hat{B}_\psi^{-1}\hat{\gamma}$$

Dove le entrate kl di \hat{B}_ψ sono

$$(\hat{B}_\psi)_{kl} = \frac{1}{b} \sum_{j=1}^b \psi_k(S_{ij})\psi_l(S_{ij})$$

e l'entrata k -esima di $\hat{\gamma}$ è

$$\hat{\gamma}_k = \frac{1}{b} \sum_{j=1}^b \psi_k(S_{ij})V_{i+1}(S_{i+1,j})$$

Notiamo che per calcolare queste quantità abbiamo bisogno di considerare solo le coppie di nodi consecutivi $(S_{ij}, S_{i+1,j})$ con $j = 1, 2, \dots, b$. Nella pratica non conosciamo esplicitamente la funzione V_{i+1} quindi anche $V_{i+1}(S_{i+1,j})$ dovrà essere sostituito con il suo valore approssimato \hat{V}_{i+1} .

1.6 Algoritmo basato sulla regressione

Dopo aver messo in luce tutti gli aspetti basilari, ci si può ora soffermare sul funzionamento dell'algoritmo. Vogliamo prezzare un'opzione bermudiana che può essere esercitata in m istanti temporali. Il sottostante è modellizzato con un moto Browniano geometrico di drift μ , volatilità σ e con valore iniziale osservabile S_0 . Supponiamo la presenza di un titolo privo di rischio con capitalizzazione continua e tasso di interesse costante annuale r . Con \hat{V}_{ij} è indicato il valore stimato dell'opzione al tempo t_i e relativo alla simulazione j .

Di seguito vengono presentati i passi dell'algoritmo:

1. Si simulano b traiettorie indipendenti del sottostante con m punti di discretizzazione $(S_{1j}, S_{2j}, \dots, S_{mj})$, $j = 1, 2, \dots, b$

2. A scadenza si pone il valore dell'opzione uguale al payoff cioè:

$$\hat{V}_{mj} = h_m(S_{mj}), \quad j = 1, 2, \dots, b$$

3. Ricorsivamente per $i = m-1, m-2, \dots, 1$:

- (a) Si sconta $\hat{V}_{i+1,j}$ al tempo t_i , cioè lo si moltiplica per il fattore sconto $e^{-r \, dt}$

- (b) Si calcolano i coefficienti di regressione

$$\hat{\beta}_i = \hat{B}_\psi^{-1} \hat{\gamma}$$

dove la regressione viene fatta tra $\psi(S_i)$ e il valore scontato calcolato nel punto precedente.

- (c) Si stima la funzione dei valori di continuazione:

$$\hat{C}_i(x) = \hat{\beta}_i \cdot \psi(x)$$

- (d) Si pone

$$\hat{V}_{ij} = \max \left\{ h_i(S_{ij}), \hat{C}_i(S_{ij}) \right\}, \quad j = 1, 2, \dots, b$$

4. Infine si calcola la stima del prezzo iniziale:

$$\hat{V}_0 = \frac{\hat{V}_{11} + \hat{V}_{12} + \dots + \hat{V}_{1b}}{b} e^{-r \, dt}$$

1.6.1 LSM

Questo algoritmo è stato introdotto da Tsitsiklis e Van Roy. Successivamente Longstaff e Schwartz lo modificarono mettendo insieme la stima dei valori di continuazione con il loro stimatore interleaving [1]. L'algoritmo è sostanzialmente lo stesso, solo il punto 3 viene sostituito dalla seguente ricorsione: Per $i = m - 1, m - 2, \dots, 1$:

- Si pone $\hat{V}_{ij} = \hat{V}_{i+1,j} e^{-r dt}$
- Si considerano poi solo le traiettorie "in the money" cioè quelle per le quali $h_i > 0$. Si stimano i coefficienti di regressione $\hat{\beta}_i$ utilizzando solo queste traiettorie.
- Si calcola la funzione dei valori di continuazione \hat{C}_i
- Considerando solo le traiettorie "in the money" si pone

$$\hat{V}_{ij} = h_i(S_{ij}), \quad \text{se } h_i(S_{ij}) \geq \hat{C}_i(S_{ij})$$

In pratica a ogni iterazione attualizzo il valore dell'opzione del nodo successivo che ho calcolato precedentemente . Tra quelli che non hanno payoff nullo modifico sono quelli dove mi conviene esercitare l'opzione.

Per quanto riguarda l'algoritmo introdotto da Tsitsiklis e Van Roy, loro stessi dimostrarono in [4] che se in ogni opportunità di esercizio la stima del valore di continuazione è esatta allora la stima del prezzo iniziale \hat{V}_0 converge al valore vero $V_0(S_0)$ quando $b \rightarrow \infty$.

Per quando riguarda invece l'algoritmo introdotto da Longstaff e Schwartz nel lavoro di Clement, Lamberton e Protter [3] viene dimostrata la convergenza al valore vero se la rappresentazione con le funzioni base è esatta. Altrimenti per $b \rightarrow \infty$ si ha un low-bias cioè:

$$E[\hat{V}_0] \leq V_0$$

Di seguito viene trattato il problema dei bias.

1.7 Bias LSM

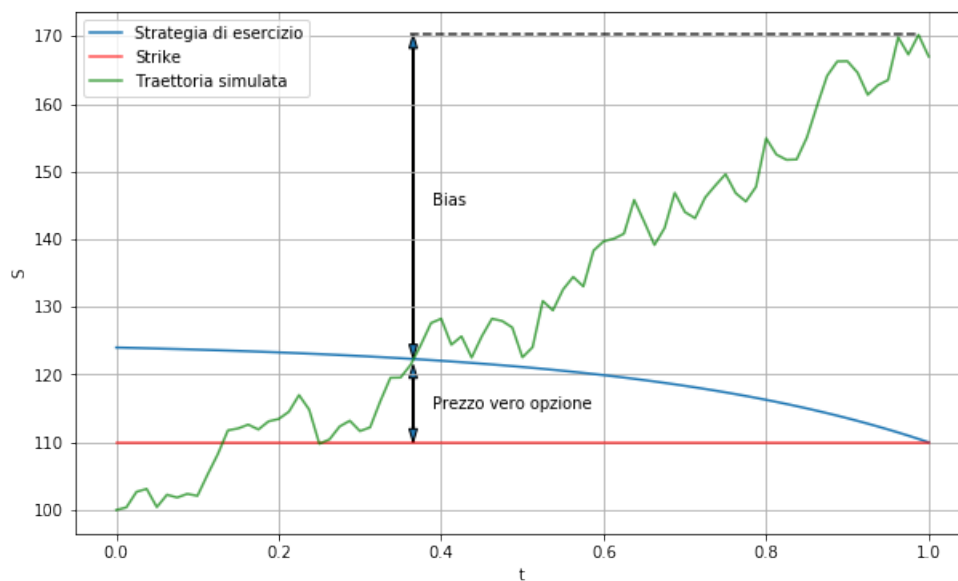
In ogni algoritmo per il pricing basato sulla simulazione si presentano di solito due tipi di bias, che si muovono in direzioni opposte: low-bias e high-bias. Anche l'LSM dunque presenta questi due tipi di bias.

Il low-bias è relativo alla decisione di esercizio subottimale. Nel caso del LSM, l'utilizzo di un numero finito di basi non può permetterci di rappresentare pienamente la funzione attesa condizionata del payoff e dunque questo ci porta a commettere un errore. La conseguenza è che la politica di esercizio si discosta da quella ottimale e si ottiene una stima minore per il prezzo

iniziale dell'opzione. Per questo motivo, questo tipo di bias viene chiamato bias subottimale. In realtà però esso risulta essere benigno. Questo perché nelle *Callable Note Markets* oltre a sgonfiare il prezzo dell'opzione, si ha che il guadagno realizzato attraverso la copertura nell'ambito della politica di esercizio non ottimale è pari al valore dell'opzione sgonfiato. In poche parole gli istituti finanziari che comprano le opzioni e agiscono anche come agenti di prezzo accettano ciò per cui pagano. L'unico aspetto negativo del bias subottimale è che gli investitori perdono scambi con concorrenti che offrono un premio di opzione più elevato.

L'high bias deriva invece dal fatto di utilizzare lo stesso insieme di simulazioni sia per la determinazione della politica di esercizio, sia per la valutazione del payoff. Questo crea una correlazione positiva tra la politica di esercizio e il payoff futuro. Infatti l'algoritmo dice di esercitare quando il payoff futuro è più basso e di continuare quando il payoff futuro è più alto.

Per capire l'errore che viene commesso prendiamo come esempio un'opzione call americana e consideriamo una traiettoria estrema che ha un payoff alla scadenza molto alto. La seguente figura può essere utile per avere un'idea più chiara di quello che succede:



L'algoritmo ci suggerisce di continuare fino alla scadenza, in quanto l'attesa del payoff futuro risulta sempre maggiore. Il problema principale sta nel fatto che la simulazione viene fatta avanti nel tempo mentre l'algoritmo lavora a ritroso. La diretta conseguenza è che quindi l'algoritmo beneficia di informazioni che nella vita reale non risultano disponibili. A differenza del bias subottimale, questo risulta essere problematico. Infatti gli investitori di solito detengono l'opzione fino alla scadenza senza copertura e pertanto sono

meno sensibili a una valutazione più accurata. Al contrario, per gli istituti finanziari la propensione al futuro è dannosa perchè gonfia erroneamente il premio dell'opzione che pagano. La presenza simultanea di questi due bias è un enorme svantaggio per lo stimatore LSM.

Si può notare il bias subottimale oltre a essere benigno non risulta eliminabile in quanto deriva da un'approssimazione propria del metodo. L'high bias invece è eliminabile. Una tecnica standard per eliminare questo bias è quella di utilizzare un ulteriore insieme indipendente di traiettorie per calcolare la politica di esercizio. In questo modo si va ad eliminare la correlazione tra la politica di esercizio e i payoff simulati. Questo approccio che utilizza quindi 2 passi elimina l'high bias ma raddoppia il costo computazionale.

Longstaff e Schwartz dimostrarono che nel caso di un singolo sottostante questo bias risulta in realtà trascurabile, mentre nei casi con più sottostanti tende a sopravvalutare l'opzione. Verrà studiato anche questo aspetto negli esperimenti numerici.

Di seguito presentiamo un altro metodo che elimina efficacemente l'high bias senza ricorrere ad ulteriori simulazioni di traiettorie. L'idea è quella di utilizzare una tecnica di convalida incrociata tipica dell'apprendimento statistico. Questa tecnica, detta anche test fuori campione, serve per valutare come i risultati di un'analisi statistica si possano generalizzare a un insieme di dati indipendenti. Viene utilizzata principalmente nei contesti dove l'obiettivo è la predizione di dati.

In un modello predittivo vengono considerati due tipi di insiemi: uno detto insieme di addestramento dove si stimano i parametri del modello e uno detto insieme test dove si valuta l'efficienza del modello. Questi due insiemi solitamente vengono ottenuti facendo una partizione del dataset. Visto in questa ottica l'high bias è un overfitting causato dall'uso dello stesso dataset per l'addestramento (cioè la stima della politica di esercizio) e per il test (cioè la valutazione delle opzioni). Usare un insieme indipendente per la politica di esercizio come nel caso del LSM con due passi corrisponde all'utilizzo del metodo hold-out che è la più semplice tecnica di test fuori campione.

Tecniche più avanzate vengono chiamate Leave- p -out e consistono nel prendere tutte le combinazioni di p elementi dal dataset per svolgere la fase di test. Noi consideriamo un caso particolare dove viene utilizzato un solo elemento per il test (cioè prendiamo $p = 1$). Indichiamo questo metodo con LOOCV. L'utilizzo di questo metodo insieme con il LSM lo chiameremo LOOLSM.

1.7.1 LOOLSM

Prima di andare a presentare il funzionamento del metodo si possono fare ulteriori considerazioni che riguardano l'implementazione del LSM:

- Al posto di utilizzare la regressione con V_{i+1} , essa si può applicare al premio o penalità di continuazione $V_{i+1} - h_i$, dunque si può stimare

$C_i - h_i$ al posto di C_i

- Se le funzioni payoff h_i vengono usate come base per la regressione le due formulazioni di sopra risultano equivalenti
- A differenza del metodo LSM nel LOOLSM utilizziamo tutti i tipi di traiettorie e non solo quelle in-the-money.
Deve essere fatto dunque un piccolo aggiustamento: Non si esercita quando $h_i(S_{ij}) = 0$ anche se $C_i(S_{ij}) < h_i(S_{ij})$

Ricordiamo che la stima del valore dell'opzione al tempo t_i e relativa alla simulazione j -esima nel caso LSM può essere riscritta in modo compatto:

$$\hat{V}_{ij} = I(C_i(S_{ij}) \geq h_i(S_{ij})) \cdot (V_{i+1,j} - h_i(S_{ij})) + h_i(S_{ij}) \quad (1.5)$$

Dove $I(\cdot)$ è la funzione indicatrice.

Per capire da dove nasce il bias applichiamo all'equazione l'attesa su tutto l'insieme delle b simulazioni condizionata al fatto che $S_{ij} = x$. Dunque la stima del prezzo dell'opzione al tempo i è data da

$$\hat{V}_i(x) = E[I(C_i(S_{ij}) \geq h_i(S_{ij})) \cdot (V_{i+1,j} - h_i(S_{ij})) | S_{ij} = x] + h_i(x)$$

Siccome nel metodo LSM, $C_i(S_{ij})$ dipende da $V_{i+1,j}$ attraverso la regressione, l'high bias deriva dalla covarianza tra la decisione di esercizio, cioè $I(C_i(S_{ij}) \geq h_i(S_{ij}))$ e il premio di continuazione $V_{ij}(S_{ij}) - h_i(S_{ij})$, che è un'informazione futura sulla traiettoria:

$$B_i(x) = \text{Cov}(I(C_i(S_{ij}) \geq h_i(S_{ij})), V_{ij}(S_{ij}) - h_i(S_{ij}) | S_{ij} = x)$$

Rimuovendo la correlazione (per esempio utilizzando un insieme indipendente di traiettorie) l'high bias viene rimosso (quindi la covarianza diventa zero) e dunque il rimanente stimatore è subottimale. Questo si può facilmente verificare:

$$\begin{aligned} \hat{V}_i(x) &= E[I(C_i(S_{ij}) \geq h_i(S_{ij})) | S_{ij} = x] \cdot E[V_{ij}(S_{ij}) - h_i(S_{ij}) | S_{ij} = x] + h_i(x) \leq \\ &\leq p(x) E[V_{i+1,j}(S_{i+1,j}) | S_{ij} = x] + (1 - p(x)) h_i(x) = \\ &= p(x) C_i(x) + (1 - p(x)) h_i(x) \leq \max(C_i(x), h_i(x)) = V_i(x) \end{aligned}$$

Dove

$$p(x) = E[I(C_i(S_{ij}) \geq h_i(S_{ij})) | S_{ij} = x]$$

Vediamo con le considerazioni fatte precedentemente come funziona il LOOLSM:

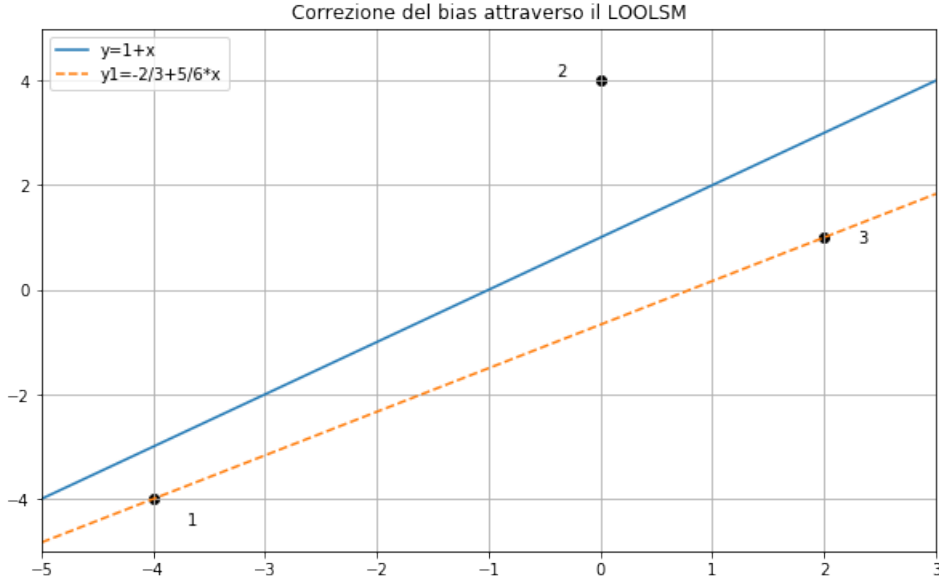
Al tempo t_i per ogni $j \in \{1, 2, \dots, b\}$ si procede come segue:

- Si stima la funzione di continuazione C_i utilizzando tutte le traiettorie tranne la j -esima

- Si stima il prezzo dell'opzione della j -esima simulazione con

$$\hat{V}_{ij} = I(C_i(S_{ij}) \geq h_i(S_{ij})) \cdot (V_{i+1,j} - h_i(S_{ij})) + h_i(S_{ij})$$

Una volta stimati tutti i prezzi al tempo t_i si passa al tempo precedente e si procede così fino al tempo 0 come nel caso LSM. Per capire meglio come funziona la regressione in questo caso possiamo considerare la seguente figura dove vengono presentate tre traiettorie al tempo i , sull'asse delle x abbiamo il valore del sottostante al tempo i e sull'asse delle y abbiamo il premio di continuazione. Vediamo cosa cambia utilizzando l'approccio LSM e quello LOOLSM in questo specifico caso in modo da comprendere a pieno l'importanza di questa modifica.



Stimando il valore di continuazione utilizzando tutte e tre le traiettorie si ottiene $y = 1 + x$, altrimenti utilizzando solo le traiettorie 1 e 3 (quindi escludendo la traiettoria 2) si ottiene $y1 = -\frac{2}{3} + \frac{5}{6}x$. Dunque secondo il metodo LSM la traiettoria 2 dovrebbe continuare, infatti $y(0) = 1 > 0$. Mentre utilizzando la correzione LOOLSM l'opzione dovrebbe essere esercitata, infatti $y1(0) = -\frac{2}{3} < 0$.

Il calcolo del prezzo dell'opzione per ogni traiettoria ad un certo tempo i richiederebbe l'applicazione della regressione b volte. In realtà il valore predetto può essere aggiustato analiticamente senza ricorrere ad ulteriori costi computazionali. Il valore di continuazione modificato può essere espresso come correzione del valore di continuazione ottenuto utilizzando tutte le traiettorie per la regressione :

$$\hat{C}'_i(S_{ij}) = \hat{C}_i(S_{ij}) - \frac{d_{ij} \cdot e_{ij}}{1 - d_{ij}} \quad j = 1, 2, \dots, b$$

Dove

$$d_{ij} = \hat{V}_{ij} - \hat{C}_i(S_{ij}) \quad e_{ij} = (H^i)_{jj} \quad H^i = \psi(S_i) \hat{B}_\psi^{-1} \psi(S_i)^T$$

Dunque l'algoritmo LOOLSM funziona come LSM, includendo tutte le traiettorie e sostituendo a \hat{C}_i il valore modificato \hat{C}'_i .

1.7.2 Convergenza bias

In questa sottosezione misuriamo l'high bias come la differenza tra la stima del prezzo iniziale fatta con LSM e quella fatta con l'LOOLSM:

$$\hat{B} = \hat{V}_0^{\text{LSM}} - \hat{V}_0^{\text{LOOLSM}}$$

Si vuole dare l'ordine di convergenza di questo bias.

Prima di presentare il teorema di convergenza abbiamo bisogno di due assunzioni tecniche:

1. Supponiamo che le funzioni payoff h_i siano continue e

$$\|h_i\|_{L^2} < +\infty \quad i = 1, 2, \dots, m$$

2. Sia $P_i(k)$ la probabilità che il valore assoluto del premio di continuazione non superi k , cioè:

$$P_i(k) = p(|C_i(S) - h_i(S)| \leq k)$$

allora

$$P_i(0) = \lim_{k \rightarrow 0^+} P_i(k) = 0$$

L'assunzione 1 deriva dal fatto che vogliamo lavorare con payoff reali che crescono in modo moderato.

L'assunzione 2 invece deriva per far fronte ad una complicazione che si può avere nell'algoritmo di pricing, che risiede nel fatto che i valori dell'opzione e i valori di continuazione possono essere arbitrariamente vicini con una probabilità non trascurabile. Questo rende sbagliata la politica di esercizio al limite e l'LSM non converge al prezzo vero.

Teorema 2. Sotto le due assunzioni valgono le seguenti affermazioni:

- Il bias \hat{B}_{ij} soddisfa:

$$\hat{B}_{ij} \sim O_p(N/b) \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, b$$

- Per ogni $\varepsilon > 0$, esiste un $r_\varepsilon > 0$ tale che il valore atteso su tutti i bias $E[\hat{B}]$ soddisfa:

$$E[\hat{B}] \leq \varepsilon + r_\varepsilon \frac{N}{b}$$

- \hat{B} converge a 0 in probabilità.

1.8 Convergenza coefficienti di regressione

Nell'algoritmo di pricing, utilizzando $N + 1$ numero di funzioni base, stimiamo i coefficienti di regressione ricorrendo ad un numero b di simulazioni. I coefficienti di regressione vengono stimati con il metodo Monte Carlo. Vogliamo studiare la convergenza di

$$E[|\beta - \hat{\beta}|^2]$$

al crescere del numero di basi e del numero di simulazioni. Questo valore atteso non è altro che l'errore quadratico medio di $\hat{\beta}$ e quindi lo indichiamo con $MSE(\hat{\beta})$.

1.8.1 Singolo Periodo

Consideriamo inizialmente un'opzione che può essere esercitata solo nei tempi $t_1 < t_2$, con t_2 che indica la scadenza dell'opzione. Dobbiamo fare la regressione di una funzione Y di S_2 sullo spazio generato da $\psi_{1k}(S_1)$ con $k = 0, 1, \dots, N$. La stima dei coefficienti è data da:

$$\beta = B_{\psi}^{-1}\gamma$$

con

$$B_{\psi} = E[\psi_1(S_1)\psi_1(S_1)^T] \quad \text{e} \quad \gamma = E[Y\psi_1(S_1)]$$

Simulando b copie indipendenti $(S_1^{(i)}, Y^{(i)})$, $i = 1, \dots, b$ si possono calcolare i coefficienti stimati:

$$\hat{\beta} = B_{\psi}^{-1}\hat{\gamma}$$

Dove $\hat{\gamma}$ è lo stimatore corretto:

$$\hat{\gamma} = \frac{1}{b} \sum_{i=1}^b Y^{(i)} \psi_1(S_1^{(i)})$$

Per semplificare l'analisi abbiamo considerato la matrice esatta B_{ψ} che di solito viene sostituita dal suo stimatore corretto. L'utilizzo del metodo con la matrice esatta da alcuni viene chiamato quasi-regressione. Indichiamo con $|x|$ la norma euclidea di un vettore x . Per una matrice A indichiamo con $\|A\|$ la norma euclidea per le matrici cioè quella uguale alla radice quadrata della somma dei quadrati delle entrate.

Ricordando che $|Ax| \leq \|A\| |x|$ e sottraendo le due equazioni abbiamo che:

$$\hat{\beta} - \beta = B_{\psi}^{-1}(\hat{\gamma} - \gamma) \tag{1.6}$$

e anche

$$B_{\psi}(\hat{\beta} - \beta) = \hat{\gamma} - \gamma \tag{1.7}$$

Passando alle norme per le due equazioni abbiamo che

$$|\hat{\beta} - \beta| = |B_{\psi}^{-1}(\hat{\gamma} - \gamma)| \leq \|B_{\psi}^{-1}\| |\hat{\gamma} - \gamma|$$

e

$$|B_{\psi}(\hat{\beta} - \beta)| = |\hat{\gamma} - \gamma| \leq \|B_{\psi}\| |\hat{\beta} - \beta|$$

Mettendo insieme le due disuguaglianze si ottiene che

$$\frac{1}{\|B_{\psi}\|} |\hat{\gamma} - \gamma| \leq |\hat{\beta} - \beta| \leq \|B_{\psi}^{-1}\| |\hat{\gamma} - \gamma| \quad (1.8)$$

Passando al quadrato (sono tutte quantità positive quindi non cambia il segno di disuguaglianza) e facendo il valore atteso abbiamo:

$$\frac{1}{\|B_{\psi}\|^2} E[|\hat{\gamma} - \gamma|^2] \leq MSE(\hat{\beta}) \leq \|B_{\psi}^{-1}\|^2 E[|\hat{\gamma} - \gamma|^2] \quad (1.9)$$

Per lo studio della convergenza di $MBE(\hat{\beta})$ dobbiamo fare le seguenti assunzioni:

(A1) $|\beta| = 1$

Siccome $MBE(\hat{\beta})$ può essere reso arbitrariamente grande moltiplicando β per una costante, adottiamo questa normalizzazione per avere risultati più significativi.

(B1) Y viene presa uguale a una combinazione lineare di funzioni base, cioè

$$Y = \sum_{k=0}^N a_k \psi_{2k}(S_2)$$

per qualche costante a_k .

In altre parole assumiamo che la rappresentazione sottoforma di regressione sia valida.

(C1) Supponiamo infine che esistano delle funzioni

$$f_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+, \quad k = 0, 1, \dots, N$$

tale che

$$E[f_k(t_2) \psi_{2k}(S_2) | S_1] = f_k(t_1) \psi_{1k}(S_1) \quad t_2 \geq t_1$$

Cioè in altre parole $\psi_{nk}(S_n)$ sono martingale contro funzioni deterministiche del tempo.

Da (B1) e (C1) abbiamo che

$$\begin{aligned}
\gamma_k &= E[Y\psi_{1k}(S_1)] = E\left[\sum_{l=0}^N a_l \psi_{2l}(S_2) \psi_{1k}(S_1)\right] = \\
&= \sum_{l=0}^N a_l E[\psi_{2l}(S_2) \psi_{1k}(S_1)] = \sum_{l=0}^N a_l E[E[\psi_{2l}(S_2) \psi_{1k}(S_1) | S_1]] = \\
&= \sum_{l=0}^N a_l E[\psi_{1k}(S_1) E[\psi_{2l}(S_2) | S_1]] = \sum_{l=0}^N a_l E\left[\frac{f_l(t_1)}{f_l(t_2)} \psi_{1k}(S_1) \psi_{1l}(S_1)\right] = \\
&= \sum_{l=0}^N a_l \frac{f_l(t_1)}{f_l(t_2)} E[\psi_{1k}(S_1) \psi_{1l}(S_1)]
\end{aligned}$$

Da (1.9) possiamo notare che la chiave per analizzare il $MBE(\hat{\beta})$ risiede nell'analisi di $E[|\hat{\gamma} - \gamma|^2]$. Siccome $\hat{\gamma}$ è una stimatore corretto di γ questo implica che $E[\hat{\gamma}] = \gamma$. Di conseguenza

$$\begin{aligned}
E[|\hat{\gamma} - \gamma|^2] &= E\left[\sum_{k=0}^N (\hat{\gamma}_k - \gamma_k)^2\right] = \sum_{k=0}^N E[(\hat{\gamma}_k - \gamma_k)^2] = \\
&= \sum_{k=0}^N Var(\hat{\gamma}_k - \gamma_k) - \sum_{k=0}^N E[\hat{\gamma}_k - \gamma_k]^2 =
\end{aligned}$$

Siccome poi $E[\gamma_k] = \gamma_k$ e $Var(\gamma_k) = 0$

$$\begin{aligned}
&= \sum_{k=0}^N Var(\hat{\gamma}_k) = \sum_{k=0}^N E[\hat{\gamma}_k^2] - \sum_{k=0}^N E[\hat{\gamma}_k]^2 = \\
&= \sum_{k=0}^N E\left[\frac{1}{b^2} \left(\sum_{i=1}^b Y^{(i)} \psi_{1k}(S_1^{(i)})\right)^2\right] - \sum_{k=0}^N \gamma_k^2 = \\
&= \frac{1}{b} \sum_{k=0}^N E[Y^2 \psi_{1k}^2(S_1)] - \frac{1}{b^2} \sum_{k=0}^N E\left[\sum_{i=1}^b \left(Y^{(i)} \psi_{1k}(S_1^{(i)})\right)^2\right] = \\
&= \frac{1}{b} \sum_{k=0}^N E[Y^2 \psi_{1k}^2(S_1)] - \frac{1}{b} \sum_{k=0}^N E\left[\frac{1}{b} \sum_{i=1}^b \left(Y^{(i)} \psi_{1k}(S_1^{(i)})\right)^2\right] = \\
&= \frac{1}{b} \sum_{k=0}^N E[Y^2 \psi_{1k}^2(S_1)] - \frac{1}{b} \sum_{k=0}^N \gamma_k^2 \tag{1.10}
\end{aligned}$$

Quindi in definitiva abbiamo che

$$E[|\hat{\gamma} - \gamma|^2] \leq \frac{1}{b} \sum_{k=0}^N E[Y^2 \psi_{1k}^2(S_1)] \quad (1.11)$$

Mettendo insieme (1.9) e (1.11) otteniamo

$$\begin{aligned} MSE(\hat{\beta}) &\leq \|B_\psi^{-1}\|^2 E[|\hat{\gamma} - \gamma|^2] \leq \|B_\psi^{-1}\|^2 \frac{1}{b} \sum_{k=0}^N E[Y^2 \psi_{1k}^2(S_1)] \leq \\ &\leq \|B_\psi^{-1}\|^2 \frac{1}{b} \sum_{l=0}^N a_l^2 \sum_{k,j=0}^N E[\psi_{2j}^2(S_2) \psi_{1k}^2(S_1)] \end{aligned} \quad (1.12)$$

Dove l'ultima disuguaglianza deriva da (B1) e dalla disuguaglianza di Cauchy-Schwarz.

Per quanto riguarda la stima dal basso definiamo un $Y^* = a_N^* \psi_{2N}(S_2)$ con a_N^* scelto in modo che il corrispondente β^* soddisfa $|\beta^*| = 1$.

Usando le proprietà del sup, la (1.9) e la (1.10) si ha che

$$\begin{aligned} \sup_{|\beta|=1} MSE(\hat{\beta}) &\geq \frac{1}{\|B_\psi\|^2} \left(\frac{1}{b} \sum_{k=0}^N E[Y^{*2} \psi_{1k}^2(S_1)] - \frac{1}{b} \sum_{k=0}^N \gamma_k^2 \right) = \\ &= \frac{1}{\|B_\psi\|^2} \left(a_N^{*2} \sum_{k=0}^N \frac{1}{b} E[\psi_{2N}^2(S_2) \psi_{1k}^2(S_2)] - \frac{1}{b} \sum_{k=0}^N \gamma_k^2 \right) \end{aligned} \quad (1.13)$$

Notiamo dalla (1.12) e la (1.13) che l'analisi della convergenza di $MSE(\hat{\beta})$ si riconduce a capire come crescono i momenti di ordine 4 della forma $E[\psi_{2j}^2(S_2) \psi_{1k}^2(S_1)]$. Il tipo di crescita dipende sia da quali funzioni base vengono utilizzate ma soprattutto dalla legge del processo S . Come abbiamo visto esso l'abbiamo modellizzato con un moto Browniano geometrico e per questo tratteremo la convergenza solo per questo caso.

Prendiamo in esame un moto Browniano geometrico con $S_0 = 1$, $\mu = 0$ e $\sigma = 1$. La dinamica di S sarà data dunque da:

$$S_t = e^{-\frac{t}{2} + W_t}$$

Le funzioni base che utilizziamo al tempo t sono date dai polinomi:

$$\psi_k(x) = x^k e^{-\frac{t}{2}k(k-1)} \quad (1.14)$$

questo perchè ci permette di ottenere le martingale:

$$\psi_k(S_t) = (S_t)^k e^{-\frac{t}{2}k(k-1)} = e^{-k\frac{t}{2} + kW_t} e^{-\frac{t}{2}k(k-1)} = e^{-k^2\frac{t}{2} + kW_t}$$

Inoltre queste funzioni soddisfano $E[\psi_k(S_t)] = 0$ e vale il seguente:

Lemma 1. Per $t_1 \leq t_2$ e $k_1, k_2 = 0, 1, \dots, N$

$$E[\psi_{k_1}(S_1)\psi_{k_2}(S_2)] = e^{k_1 k_2 t_1}$$

$$E[\psi_{k_1}^2(S_1)\psi_{k_2}^2(S_2)] = e^{k_1^2 t_1 + k_2^2 t_2 + 4k_1 k_2 t_1}$$

e $E[\psi_{k_1}^2(S_1)\psi_{k_2}^2(S_2)]$ sono strettamente crescenti in k_1 e k_2 .

Dimostrazione. Utilizzando le proprietà dell'attesa condizionata, la proprietà di martingala e ricorrendo alla funzione generatrice dei momenti di W_{t_1} abbiamo che:

$$\begin{aligned} E[\psi_{k_1}(S_1)\psi_{k_2}(S_2)] &= E[E[\psi_{k_1}(S_1)\psi_{k_2}(S_2)|W_{t_1}]] = \\ &= E[\psi_{k_1}(S_1)E[\psi_{k_2}(S_2)|W_{t_1}]] = E[\psi_{k_1}(S_1)\psi_{k_2}(S_1)] = \\ &= E[e^{-(k_1^2 + k_2^2)\frac{t_1}{2} + (k_1 + k_2)W_{t_1}}] = e^{-(k_1^2 + k_2^2)\frac{t_1}{2}} E[e^{(k_1 + k_2)W_{t_1}}] = \\ &= e^{-(k_1^2 + k_2^2)\frac{t_1}{2}} e^{(k_1 + k_2)^2 \frac{t_1}{2}} = e^{k_1 k_2 t_1} \end{aligned}$$

Procedendo in modo simile dimostriamo anche la seconda uguaglianza:

$$\begin{aligned} E[\psi_{k_1}^2(S_1)\psi_{k_2}^2(S_2)] &= E[E[\psi_{k_1}^2(S_1)\psi_{k_2}^2(S_2)|W_{t_1}]] = \\ &= E[\psi_{k_1}^2(S_1)E[\psi_{k_2}^2(S_2)|W_{t_1}]] = E[\psi_{k_1}^2(S_1)E[e^{k_2^2 t_2} \psi_{2k_2}(S_2)|W_{t_1}]] \\ &= E[e^{k_2^2 t_2} \psi_{k_1}^2(S_1)\psi_{2k_2}(S_1)] = e^{k_2^2 t_2} E[e^{-(2k_1^2 + 4k_2^2)\frac{t_1}{2} + 2(k_1 + k_2)W_{t_1}}] = \\ &= e^{k_2^2 t_2 - (k_1^2 + 2k_2^2)t_1} E[e^{2(k_1 + k_2)W_{t_1}}] = e^{k_2^2 t_2 - (k_1^2 + 2k_2^2)t_1 + 2t_1(k_1 + k_2)^2} \\ &= e^{k_1^2 t_1 + k_2^2 t_2 + 4k_1 k_2 t_1} \end{aligned}$$

La stretta crescita in k_1 e k_2 è una semplice conseguenza. \square

Diamo ora anche il seguente lemma che ci è utile per il nostro studio senza però darne la dimostrazione.

Lemma 2. Per la nostra matrice $B_\psi(t)$ vale la seguente stima

$$\|B_\psi(t)\|^2 \leq (N+1)^2 e^{2N^2 t}$$

e inoltre la sua inversa soddisfa

$$\|B_\psi^{-1}(t)\| \leq K(t)^{-1} N(N+1) \left(\frac{e^t}{e^t - 1} \right)^N$$

Dove $K(t) = e^{\frac{-2e}{(e^t - 1)^2}}$.

Utilizzando questi 2 lemmi possiamo dimostrare il seguente risultato di convergenza.

Teorema 3. Consideriamo le ψ_k definite come in (1.14).

Se

$$N = \sqrt{\frac{(1 - \delta) \log b}{5t_1 + t_2}}$$

per qualche $\delta > 0$ allora:

$$\lim_{b \rightarrow \infty} \sup_{|\beta|=1} MSE(\hat{\beta}) = 0$$

Se invece

$$N = \sqrt{\frac{(1 + \delta) \log b}{3t_1 + t_2}}$$

per qualche $\delta > 0$ allora:

$$\lim_{b \rightarrow \infty} \sup_{|\beta|=1} MSE(\hat{\beta}) = +\infty$$

Dimostrazione. Dall'assunzione (B1) abbiamo che

$$Y = \sum_{k=0}^N a_k \psi_{2k}(S_2)$$

Vogliamo calcolare il valore atteso di Y a S_1 :

$$\begin{aligned} E[Y|S_1] &= E \left[\sum_{k=0}^N a_k \psi_k(S_2) | S_1 \right] = \sum_{k=0}^N a_k E[\psi_k(S_2) | S_1] = \\ &= \sum_{k=0}^N a_k \psi_k(S_1) \end{aligned}$$

Dove nell'ultimo passaggio abbiamo sfruttato la proprietà di martingala di $\psi_k(S_t)$. Questo vuol dire che $\beta_k = a_k$, $k = 0, 1, \dots, N$. In questo caso quindi $|(a_0, a_1, \dots, a_N)| = |\beta| = 1$.

Utilizzando i due lemmi e la (1.12) abbiamo che :

$$\begin{aligned} \sup_{|\beta|=1} MSE(\hat{\beta}) &\leq \sup_{|\beta|=1} \|B_\psi^{-1}\|^2 \frac{1}{b} E \left[\sum_{k=0}^N \psi_k^2(S_2) \psi_N^2(S_1) \right] \leq \\ &\leq \|B_\psi^{-1}\|^2 \frac{1}{b} (N+1) E[\psi_N^2(S_2) \psi_N^2(S_1)] \leq \\ &\leq K^{-2}(t_1) N^2 \frac{(N+1)^3}{b} \left(\frac{e^{t_1}}{e^{t_1} - 1} \right)^{2N} e^{5N^2 t_1 + N^2 t_2} \end{aligned} \quad (1.15)$$

Ora prendiamo $N = \sqrt{\frac{(1-\delta)\log b}{5t_1+t_2}}$ e applicando il logaritmo otteniamo:

$$\begin{aligned}
& \log \left[K^{-2}(t_1) N^2 \frac{(N+1)^3}{b} \left(\frac{e^{t_1}}{e^{t_1}-1} \right)^{2N} e^{5N^2 t_1 + N^2 t_2} \right] = \\
& = \log \left[K^{-2}(t_1) N^2 \frac{(N+1)^3}{b} \left(\frac{e^{t_1}}{e^{t_1}-1} \right)^{2N} e^{(1-\delta)\log b} \right] = \\
& = \log \left[K^{-2}(t_1) N^2 (N+1)^3 \left(\frac{e^{t_1}}{e^{t_1}-1} \right)^{2N} b^{-\delta} \right] = \\
& = -\delta \log b + \log \left[K^{-2}(t_1) N^2 (N+1)^3 \left(\frac{e^{t_1}}{e^{t_1}-1} \right)^{2N} \right] = \\
& = -\delta \log b + o(\log b) \rightarrow -\infty \quad \text{per } b \rightarrow \infty
\end{aligned}$$

Questo vuol dire che l'argomento del logaritmo va a 0 quando $b \rightarrow \infty$. Dalla (1.15) e dal fatto che $MSE(\hat{\beta}) \geq 0$ (per il teorema del confronto) si ha che

$$\lim_{b \rightarrow \infty} \sup_{|\beta|=1} MSE(\hat{\beta}) = 0$$

Questo conclude la prima parte della dimostrazione.

Prendiamo ora la (1.13) dove scegliamo $Y^* = \psi_N(S_2) = e^{-N^2 \frac{t_2}{2} + NW_{t_2}}$. Utilizzando i due lemmi otteniamo:

$$\begin{aligned}
\sup_{|\beta|=1} MSE(\hat{\beta}) & \geq \frac{1}{||B_\psi||^2} \left(\sum_{k=0}^N \frac{1}{b} E[Y^* \psi_k^2(S_1)] - \frac{1}{b} \sum_{k=0}^N (\gamma_k^*)^2 \right) \geq \\
& \geq \frac{1}{b(N+1)^2 e^{2N^2 t_1}} (e^{5N^2 t_1 + N^2 t_2} - e^{2N^2 t_1}) = \\
& = \frac{1}{b(N+1)^2} (e^{3N^2 t_1 + N^2 t_2} - 1)
\end{aligned}$$

Posto ora $N = \sqrt{\frac{(1+\delta)\log b}{3t_1+t_2}}$ e applicando il logaritmo si ha:

$$\log \left[\frac{1}{b(N+1)^2} (b^{1+\delta} - 1) \right] = \delta \log b + o(\log b) \rightarrow \infty \quad \text{per } b \rightarrow \infty$$

Questo conclude la dimostrazione. \square

Il risultato di questo teorema ci mostra che N deve essere più piccolo dell'ordine di $\sqrt{\log b}$ e conseguentemente b deve essere più grande dell'ordine di e^{N^2} .

1.8.2 Multiperiodo

Passiamo ora a considerare il caso generale, cioè quello con m opportunità di esercizio. Anche per questo caso facciamo le seguenti assunzioni :

(A2) $E[\psi_{ik}^2(S_i)]$ e $E[\psi_{ik}^4(S_i)]$ sono crescenti in i e k .

(B2) $|\beta_{m-1}| = 1$

Come nel caso precedente abbiamo bisogno di questa normalizzazione per avere risultati significativi.

(C2)

$$E[h_i^4(S_i)] \leq \left(\frac{t_i}{t_{i-1}}\right)^{2N} E[\psi_{iN}^4(S_i)] \quad i = 0, 1, \dots, m$$

Quindi imponiamo una condizione anche sulle funzioni payoff h_i .

Supponendo che gli S_i abbiamo una densità g_i definiamo la norma L^2 pesata su una funzione $F : \mathbb{R} \rightarrow \mathbb{R}$:

$$\|F\|_i = \sqrt{\int F(x)^2 g_i(x) dx}$$

Indicando con C_i e \hat{C}_i rispettivamente la funzione di continuazione e la stima con la regressione che deriva dal LSM, andiamo ad analizzare l'errore quadratico medio in norma L^2 . Cioè vogliamo calcolare $E[\|\hat{C}_i - C_i\|_i^2]$.

Diamo delle notazioni aggiuntive che ci servono per il nostro studio:

$$c = \max_{i=1,2,\dots,m-1} \frac{t_{i+1}}{t_i} \quad B_N = \max_{i=1,2,\dots,m-1} \|B_\psi^{-1}(t_i)\|$$

$$H_N = \max\{c^N, B_N^2(N+1)\} \quad A_N = (N+1)H_N E[\psi_{mN}^2(S_m)]$$

Abbiamo il seguente importante risultato:

Teorema 4. Supponiamo che valgano le assunzioni (A0), (A2)-(C2).

Allora:

$$E[\|\hat{C}_i - C_i\|_i^2] \leq (2^{m-i} - 1) \frac{(N+1)^2}{b} B_N A_N^{m-i} (E[\psi_{mN}^2(S_m)])^2 (1 + o(1))$$

Per dimostrare il teorema abbiamo del seguente lemma di cui però non diamo la dimostrazione.

Lemma 3. Supponiamo che valgano le assunzioni (A0), (A2)-(C2).

Allora:

$$|\gamma_{m-i}|^2 \leq (2H_N E[\psi_{mN}^4(S_m)])^{i-1} (N+1)^{i+1} (E[\psi_{mN}^2(S_m)])^2 (1 + o(1))$$

$$E[|\hat{C}_i - C_i|^2] \leq B_N \sum_{l=1}^{m-i} A_N^{m-i-l} E[|\tilde{\gamma}_{m-l} - \gamma_{m-l}|^2]$$

Per $i = 1, 2, \dots, m-1$.

Con $\tilde{\gamma}_i$ indichiamo la stima di γ_i se noi conosciamo esplicitamente V_{i+1} . Ricordiamo che invece $\hat{\gamma}_i$ è la stima utilizzando \hat{V}_{i+1} .

Andiamo ora a provare il nostro teorema.

Dimostrazione. Siccome ogni $\tilde{\gamma}_{ik}$ è uno stimatore corretto del corrispondente γ_{ik} allora $E[(\tilde{\gamma}_{ik} - \gamma_{ik})^2] = \text{Var}(\tilde{\gamma}_{ik}) \leq E[\tilde{\gamma}_{ik}^2]$. Dunque

$$\begin{aligned} E[|\tilde{\gamma}_{m-i} - \gamma_{m-i}|^2] &= \sum_{k=0}^N E[(\tilde{\gamma}_{m-i,k} - \gamma_{m-i,k})^2] \leq \\ &\leq \sum_{k=0}^N \frac{1}{b} E[\psi_{m-i,k}^2(S_{m-i}) \max\{h_{m-i+1}^2(S_{m-i+1}), C_{m-i+1}^2(S_{m-i+1})\}] \leq \\ &\leq \sum_{k=0}^N \frac{1}{b} E[\psi_{m-i,k}^2(S_{m-i}) (h_{m-i+1}^2(S_{m-i+1}) + C_{m-i+1}^2(S_{m-i+1}))] \leq \\ &\leq \sum_{k=0}^N \frac{1}{b} E[\psi_{m-i,k}^2(S_{m-i}) (h_{m-i+1}^2(S_{m-i+1}) + \|B_\psi^{-1}(t_{m-i})\|^2 |\gamma_{m-i}|^2 |\psi_{m-i+1}(S_{m-i+1})|^2)] \end{aligned}$$

Dove abbiamo usato il fatto che per ogni $x \in \mathbb{R}$

$$C_i^2(x) = (\psi_i^T(x) B_\psi^{-1}(t_i) \gamma_i)^2 \leq |\psi_i(x)|^2 \|B_\psi^{-1}(t_i)\|^2 |\gamma_i|^2$$

Ora spezziamo la sommatoria e facciamo due stime separate per i due termini.

$$\begin{aligned} &\sum_{k=0}^N \frac{1}{b} E[\psi_{m-i,k}^2(S_{m-i}) h_{m-i+1}^2(S_{m-i+1})] \leq \\ &\leq \frac{N+1}{b} \sqrt{E[\psi_{mN}^4(S_m)] E[h_{m-i+1}^4(S_{m-i+1})]} \leq \\ &\leq \frac{N+1}{b} H_N E[\psi_{mN}(S_m)] \end{aligned}$$

Dove nel primo passaggio abbiamo usato la disuguaglianza di Cauchy-Schwartz e (A2) e nel secondo passaggio abbiamo usato (C2). Per il secondo termine della sommatoria abbiamo

$$\begin{aligned} &\sum_{k=0}^N \frac{1}{b} E[\psi_{m-i,k}^2(S_{m-i}) \|B_\psi^{-1}(t_{m-i})\|^2 |\gamma_{m-i}|^2 |\psi_{m-i+1}(S_{m-i+1})|^2] \leq \\ &\leq \frac{N+1}{b} H_N E[\psi_{mN}^4(S_m)] |\gamma_{m-i}|^2 \end{aligned}$$

Dove abbiamo usato ancora la disuguaglianza di Cauchy-Schwartz e (A2). Combinando la prima stima del lemma 3 e le tre stime che abbiamo calcolato otteniamo che

$$\begin{aligned} E[|\tilde{\gamma}_{m-i} - \gamma_{m-i}|^2] &\leq \\ &\leq \frac{(N+1)^{i+2}}{b} 2^{i-1} (H_N E[\psi_{mN}^4(S_m)])^i (E[\psi_{mN}^2(S_m)])^2 (1 + o(1)) = \\ &= \frac{2^{i-1} (K+1)^2}{b} A_N^i (E[\psi_{mN}^2(S_m)])^2 (1 + o(1)) \end{aligned}$$

Da questa stima e dalla seconda stima derivante dal lemma 3 otteniamo:

$$\begin{aligned} E[||\hat{C}_i - C_i||_i^2] &\leq B_N \left(\sum_{l=1}^{m-i} A_N^{m-i-l} E[|\tilde{\gamma}_{m-l} - \gamma_{m-l}|^2] \right) \leq \\ &\leq B_N \frac{(N+1)^2}{b} A_N^{m-i} (E[\psi_{mN}^2(S_m)])^2 \cdot (1 + 2 + \dots + 2^{m-i-1}) (1 + o(1)) = \\ &= (2^{m-1} - 1) B_N \frac{(N+1)^2}{b} A_N^{m-i} (E[\psi_{mN}^2(S_m)])^2 (1 + o(1)) \end{aligned}$$

Che conclude la dimostrazione. \square

Come nel caso di un singolo periodo, prendiamo S modellizzato con un moto Browniano geometrico e usiamo le funzioni base (1.14). In questo caso abbiamo che

$$B_N = \max_i ||B_\psi^{-1}(t_i)|| < \max_i C^{-1}(t_i) \left(\frac{e^{t_i}}{e^{t_i} - 1} \right)^{N-1} < e^{\frac{2e}{e^{t_1}-1}} \left(\frac{e^{t_1}}{e^{t_1} - 1} \right)^{N-1}$$

La prima disuguaglianza deriva dal lemma 1 e il secondo dal fatto che $C^{-1}(t_i)$ e $\left(\frac{e^{t_1}}{e^{t_1}-1} \right)^{N-1}$ realizzano il loro massimo per $i = 1$.

Dai lemmi 1 e 2 abbiamo che

$$E[\psi_{mN}^4(S_m)] = e^{6N^2 t_m} \quad \text{e} \quad E[\psi_{mN}^2(S_m)] = e^{N^2 t_m}$$

Facendo queste sostituzioni in A_N e nella stima del teorema otteniamo che

$$\begin{aligned} E[||\hat{C}_i - C_i||_i^2] &< \\ &< (2^{m-i} - 1) \frac{(N+1)^{m-i+2}}{b} B_N H_N^{m-i} e^{6(m-i)N^2 t_m + 2N^2 t_m} (1 + o(1)) \end{aligned}$$

$(N+1)^2$ è trascurabile rispetto al fattore esponenziale. B_N e H_N crescono esponenzialmente in N ma i loro esponenti sono lineari in N e quindi il fattore dominante risulta essere quelli quadratici in N . Quindi B_N e H_N sono trascurabili per N abbastanza grande. Se poniamo

$$N = \sqrt{\frac{(1-\delta) \log b}{(6(m-i) + 2)t_m}}$$

per qualche $\delta > 0$ allora

$$\lim_{b \rightarrow \infty} \sup_{|\beta_{m-1}|=1} E[||\hat{C}_i - C_i||_i^2] = 0$$

Inoltre dal teorema 2 si ha che se

$$N = \sqrt{\frac{(1 + \delta) \log b}{3t_m + t_{m-1}}}$$

per qualche $\delta > 0$ allora:

$$\lim_{b \rightarrow \infty} \sup_{|\beta_{m-1}|=1} E[||\hat{C}_{m-1} - C_{m-1}||_{m-1}^2] = \infty$$

Dunque il tasso critico di N è $O(\sqrt{\log b})$ come nel caso del problema con un singolo periodo.

1.9 Opzioni Americane

Fino adesso abbiamo trattato il calcolo del valore iniziale di un'opzione sempre dal punto di vista numerico. In questa sezione vogliamo dare anche un'idea di approccio analitico che può essere usato per risolvere il problema.

Definizione. Un problema con ostacolo è un problema del tipo

$$\begin{cases} \max\{Lu, \varphi - u\} = 0, & \text{in } \mathcal{S}_T = (0, T) \times \mathbb{R}^d \\ u(0, \cdot) = \varphi & \text{in } \mathbb{R}^d \end{cases}$$

Dove L è un operatore parabolico della forma:

$$Lu := \frac{1}{2} \sum_{i,j=1}^d c_{ij} \partial_{x_i x_j} u + \sum_{i=1}^d b_i \partial_{x_i} u - au - \partial_t u$$

con $(t, x) \in \mathbb{R} \times \mathbb{R}^d$, (c_{ij}) è una matrice simmetrica. Inoltre assumiamo che c_{ij}, b_i, a sono funzioni di (t, x) limitate e Hölderiane. φ è una funzione localmente Lipschitziana e convessa in senso debole.

La soluzione di un problema di questo tipo è una funzione u tale che :

- i) $Lu \leq 0$
- ii) $u \geq \varphi$
- iii) $Lu = 0$ nel caso in cui $u > \varphi$
- iv) $u(0, x) = \varphi(0, x), \quad x \in \mathbb{R}^d$

Vediamo che il prezzo di un'opzione americana con payoff φ si può esprimere in termini della soluzione u di un problema con ostacolo.

Consideriamo un sottostante con la solita dinamica:

$$dS_t = (r - q)S_t dt + \sigma S_t dW_t$$

Vediamo ora il risultato che mette in relazione il problema con ostacolo parabolico con il corrispondente problema per l'operatore differenziale di Black&Scholes:

$$L_{BS}f(t, S) := \frac{\sigma^2 S^2}{2} \partial_{SS} f(t, S) + (r - q)S \partial_S f(t, S) + \partial_t f(t, S) - rf(t, S)$$

Teorema 5. Esiste ed è unica la soluzione forte $f \in C([0, T] \times \mathbb{R}^+)$ per il problema con ostacolo:

$$\begin{cases} \max\{L_{BS}f, \psi - f\} = 0, & \text{in } (0, T) \times \mathbb{R}^+ \\ f(T, \cdot) = \psi(T, \cdot) & \text{in } \mathbb{R}^+ \end{cases}$$

Dove $\psi(t, S_t)$ rappresenta il premio ottenuto esercitando l'opzione al tempo t .

Inoltre il problema soddisfa le seguenti proprietà:

- i) Per ogni $(t, y) \in [0, T] \times \mathbb{R}^+$ si ha

$$f(t, y) = \sup_{\tau \in \mathcal{T}, \tau \in [t, T]} E^Q \left[e^{-r(\tau-t)} \psi(\tau, S_\tau^{t,y}) \right]$$

Dove $S^{t,y}$ è un moto Browniano geometrico con condizione iniziale $S_t = y$.

- ii) f ammette derivata parziale prima rispetto ad S in senso classico e vale

$$\partial_S f \in C \cap L^\infty((0, T) \times \mathbb{R}^+)$$

La striscia $\mathcal{S}_T = (0, T) \times \mathbb{R}^+$ è suddivisa in due parti:

- i) La *regione di esercizio* dove $f = \psi$
ii) La *regione di continuazione* dove $f > \psi$ e $L_{BS}f = 0$.

1.10 Metodo Monte Carlo

Il Metodo Monte Carlo è un metodo numerico basato sulla simulazione che permette di affrontare problemi che analiticamente risultano di difficile risoluzione.

1.10.1 Cenni storici

Nel 1777 il matematico francese Georges-Louis Leclerc de Buffon pose le basi teoriche di questo metodo proponendo e risolvendo il problema noto come "L'ago di Buffon". Questo esperimento consisteva nel lanciare un ago su un piano intersecato da rette parallele tutte poste alla stessa distanza. Conoscute la distanza tra le rette e la lunghezza dell'ago attraverso delle prove ripetute si poteva dare una stima del π contando il numero di volte che l'ago intersecava le rette.

Agli inizi dell'1800 Laplace diede delle robuste basi matematiche per lo sviluppo del metodo.

Negli anni '30 Enrico Fermi utilizzò la simulazione numerica per studiare il moto di diffusione dei neutroni. Negli anni '40 Stan Ulam cercò di calcolare la probabilità di vittoria al solitario in base alla disposizione iniziale delle carte. L'utilizzo del calcolo combinatorio per la risoluzione del problema risultava difficoltoso e per questo decise di testare numerose disposizioni iniziali e contare il numero delle partite vittoriose. Ulam si rese conto che un approccio di questo tipo si prestava ad essere eseguito con l'ausilio dei calcolatori.

Lo stesso propose a John Von Neumann di utilizzare questo metodo di simulazione numerica per studiare la dinamica delle esplosioni nucleari nell'ambito del progetto Manhattan nei laboratori di Los Alamos. John rimase affascinato dall'idea di usare la simulazione numerica e spinse affinché il metodo venisse inserito nel progetto. Nick Metropolis progettò delle strutture hardware ottimizzate per il metodo e gli diede il nome Monte Carlo anche come copertura, in quanto il progetto era top secret.

Nel 1949 Ulam e Metropolis pubblicarono il primo articolo sul metodo Monte Carlo con l'aiuto di Fermi e Von Neumann. Questo viene seguito da altri lavori che presentavano come il metodo potesse essere utilizzato per risolvere alcuni problemi della Fisica.

Il metodo così iniziò a diffondersi anche in ambito statistico, fino ad arrivare ai giorni d'oggi dove il metodo è utilizzato in Matematica, Fisica, Statistica, Ingegneria e Finanza.

1.10.2 Principi matematici

I metodi Monte Carlo sono una classe di algoritmi che si basano sull'analogia tra probabilità e volume. L'idea principale di questi metodi è quella di voler risolvere problemi deterministici ricorrendo alla casualità. Attraverso la teoria della misura viene formalizzata l'idea intuitiva di probabilità, se consideriamo un evento come un insieme di differenti configurazioni, la probabilità di questo evento è data dalla sua misura o dal suo volume. Questo è vero per tutti i possibili eventi.

I Metodi Monte Carlo calcolano la misura di un insieme e interpretano que-

sto volume come una probabilità.

Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e sia $X : \Omega \rightarrow \mathbb{R}^d$ una variabile aleatoria su suddetto spazio. Consideriamo $f \in m\mathcal{B}$, una funzione Borel-misurabile, e supponiamo che $f(X) \in L^2(\Omega, P)$. Vogliamo calcolare il valore atteso $E[f(X)]$.

Dalla legge dei grandi numeri abbiamo che:

$$E[f(X)] \approx \frac{1}{n} \sum_{i=1}^n f(X^{(i)}), \quad n \gg 1$$

Dove $X^{(i)}$ con $i = 1, \dots, n$ sono realizzazioni indipendenti della variabile casuale X . Conoscendo la distribuzione di X possono essere simulati n suoi valori e si può approssimare il valore atteso con la somma finita vista sopra.

Capitolo 2

Esempi Numerici

In questo capitolo vengono considerati diversi esperimenti numerici che confermano i risultati ottenuti dal punto di vista teorico. Oltre a valutare la opzioni con il metodo LSM, vengono fatti dei confronti tra LOOLSM e LSM per quanto riguarda i bias, il costo computazionale e l'errore relativo. Inoltre sono stati considerati degli esempi sulla convergenza dei coefficienti di regressione.

2.1 Call Americana con un sottostante

Consideriamo un'opzione americana di tipo Call, dunque il payoff è dato da

$$h_i = (S_i - K)^+ \quad i = 1, 2, \dots, m$$

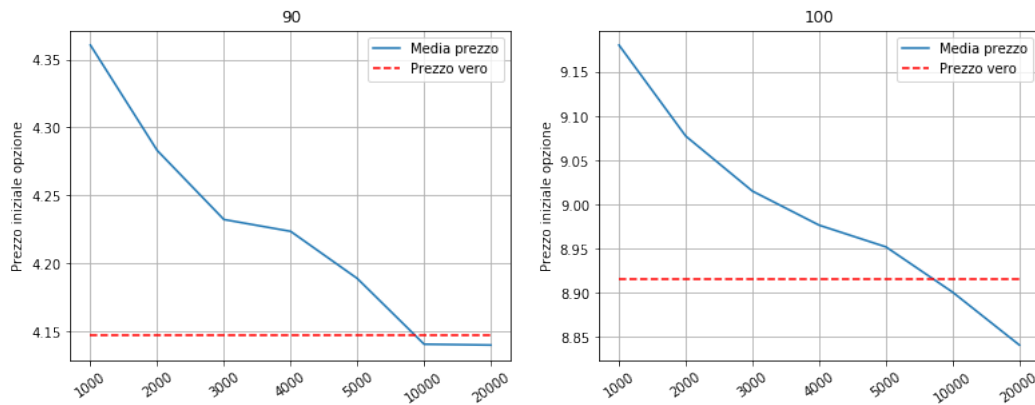
K indica lo strike ed è stato mantenuto costante uguale a 100 in tutti gli esperimenti. L'opzione americana può essere esercitata nell'intervallo $[0, 1]$. È stato valutato il prezzo iniziale dell'opzione sia utilizzando LSM sia utilizzando l'algoritmo base di regressione di Tsitsiklis e Van Roy. Gli esperimenti sono stati fatti variando i seguenti parametri: la volatilità, il valore iniziale del sottostante e il numero di discretizzazioni. Il valore del sottostante è stato modellizzato con un moto Browniano geometrico con drift $r = 0.02$, parametro che abbiamo mantenuto costante. Sono state usate 20000 simulazioni per il prezzaggio e per la regressione sono state usate come basi $1, x, x^2, x^3$. Di seguito viene riportata una tabella che riassume i risultati ottenuti. Nell'ultima colonna è riportato la stima del prezzo utilizzando gli alberi binomiali dove sono state considerate 10000 opportunità di esercizio.

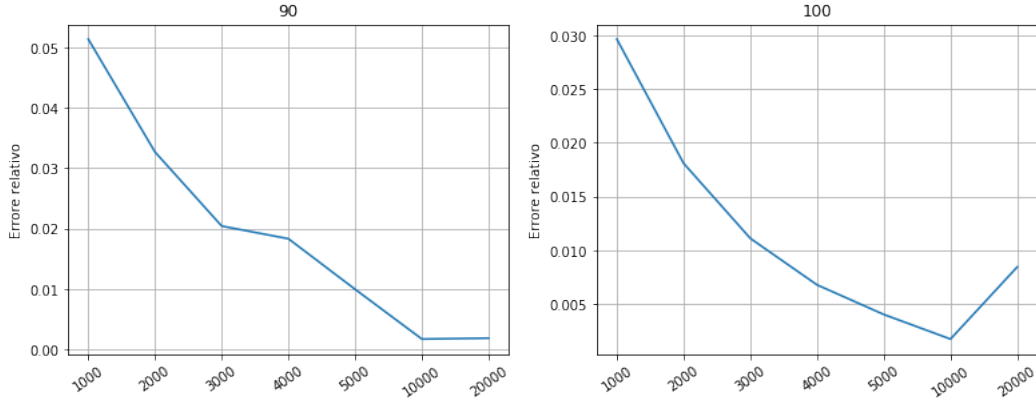
S_0	σ	m	LSM	Regressione	BT
80	0.2	50	1.4285	2.1366	1.4273
		150	1.4110	2.3146	
	0.4	50	6.7732	8.8611	6.8501
		150	6.7122	9.2813	
90	0.2	50	4.1662	5.403	4.1477
		150	4.0911	5.6538	
	0.4	50	11.2388	13.2257	11.2258
		150	11.2029	13.8154	
100	0.2	50	8.9515	10.0264	8.9158
		150	8.7800	10.3122	
	0.4	50	16.6735	18.9889	16.7040
		150	16.6212	19.8469	
110	0.2	50	15.6533	16.6877	15.6089
		150	15.6096	17.1044	
	0.4	50	23.2287	26.0680	23.1479
		150	22.9004	27.1606	
120	0.2	50	23.7881	25.0155	23.7422
		150	23.6512	25.4982	
	0.4	50	30.4065	34.0206	30.3971
		150	30.1786	35.2802	

Fissando i seguenti parametri:

$$\sigma = 0.2 \quad r = 0.02 \quad m = 300$$

Per $S_0 = 90, 100$ è stato replicato 150 volte l'algoritmo con un numero crescente di simulazioni $b = 1000, 2000, 3000, 4000, 5000, 10000, 20000$. Per ogni b fissato è stata fatta la media sulle 150 repliche del prezzo iniziale stimato ed è stato calcolato l'errore relativo rispetto al "prezzo vero" ottenuto con gli alberi binomiali. Di seguito sono riportati i grafici della media dei prezzi e degli errori relativi per i due valori iniziali del sottostante.





Possiamo notare che al crescere di b lo stimatore diventa un low-bias. Questo conferma il fatto che l'high bias è trascurabile quando si considera un solo sottostante. Inoltre possiamo notare che per b abbastanza grande l'errore relativo risulta minore del 1% per entrambi i valori iniziali dei sottostanti.

2.2 Esempi numerici con più sottostanti

In questa sezione verranno considerati degli esperimenti numerici fatti su opzioni che dipendono da più sottostanti. Sono stati presi in esame casi in cui i sottostanti risultano indipendenti e casi in cui i sottostanti sono correlati.

2.2.1 Massimo di due sottostanti indipendenti

Iniziamo con il studiare un'opzione call sul massimo di due sottostanti indipendenti. Dunque il payoff è dato da:

$$h(S_1(t), S_2(t)) = (\max(S_1(t), S_2(t)) - K)^+$$

Con strike $K = 100$ che è stato mantenuto costante in tutti gli esperimenti. Inoltre

$$S_1 \sim MBG(r - \delta, \sigma) \quad S_2 \sim MBG(r - \delta, \sigma)$$

Con i parametri mantenuti costanti

$$r = 0.05 \quad \delta = 0.1 \quad \sigma = 0.2$$

È stata considerata un'opzione bermudiana con scadenza al tempo $T = 3$ anni, esercitabile nei tempi $t_i = i/3$, $i = 1, 2, \dots, 9$. Per gli esperimenti sono state simulate 4000 traiettorie per ogni sottostante. Sono stati fatti esperimenti utilizzando anche basi diverse. Ricorrendo alla stima del prezzo dell'opzione con un modello binomiale bidimensionale per i valori iniziali

$S_i(0) = 90, 100, 110$, $i = 1, 2$, si ottiene rispettivamente 8.08, 13.90, 21.34. Di seguito ci sono le tabelle con le stime utilizzando l'algoritmo base di Tsitsiklis e Van Roy e il LSM, inoltre sono riportate le funzioni base utilizzate nella regressione.

$$S_1(0) = S_2(0) = 90$$

Funzioni base	LSM	Regressione
$1, S_i, S_i^2, S_i^3$	8.035	9.373
$1, S_i, S_i^2, S_i^3, S_1 S_2$	8.024	9.284
$1, S_i, S_i^2, S_i^3, S_1 S_2, \max(S_1, S_2)$	8.003	9.415
$1, S_i, S_i^2, S_i^3, S_1 S_2, S_1^2 S_2, S_1 S_2^2$	8.072	9.397

$$S_1(0) = S_2(0) = 100$$

Funzioni base	LSM	Regressione
$1, S_i, S_i^2, S_i^3$	13.842	15.058
$1, S_i, S_i^2, S_i^3, S_1 S_2$	13.988	15.208
$1, S_i, S_i^2, S_i^3, S_1 S_2, \max(S_1, S_2)$	13.712	15.064
$1, S_i, S_i^2, S_i^3, S_1 S_2, S_1^2 S_2, S_1 S_2^2$	13.860	14.843

$$S_1(0) = S_2(0) = 110$$

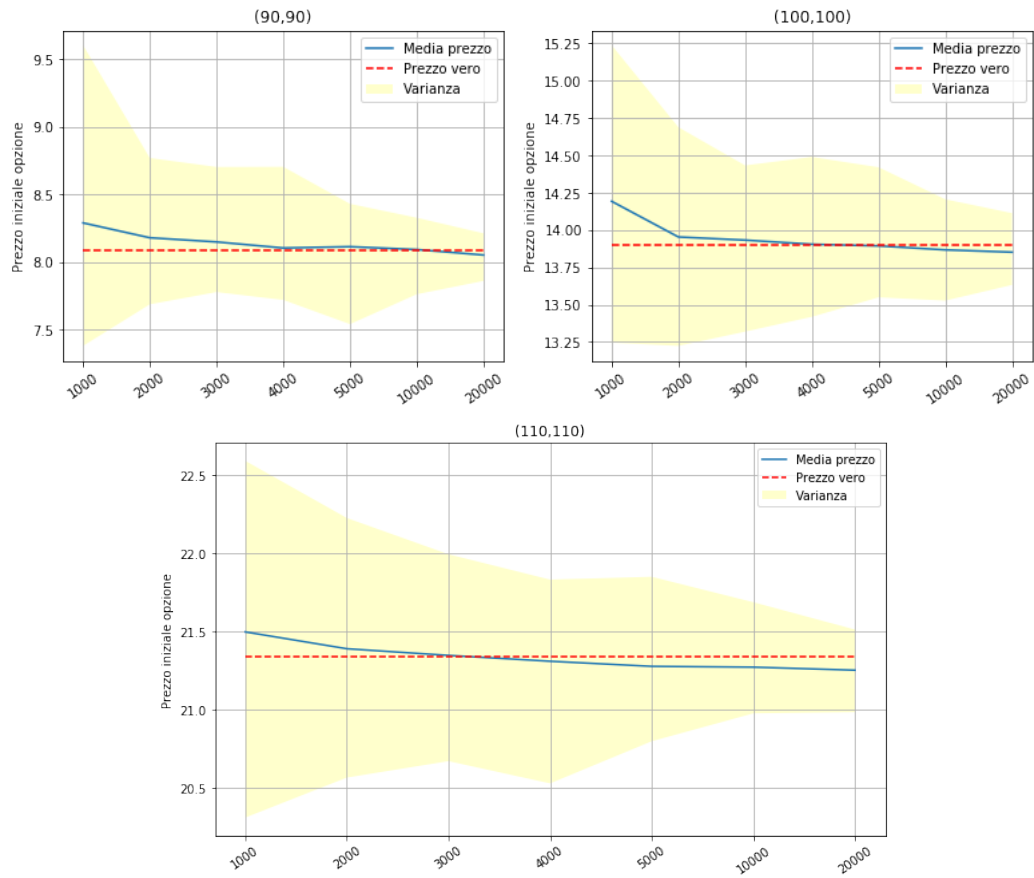
Funzioni base	LSM	Regressione
$1, S_i, S_i^2, S_i^3$	21.611	22.236
$1, S_i, S_i^2, S_i^3, S_1 S_2$	21.336	22.616
$1, S_i, S_i^2, S_i^3, S_1 S_2, \max(S_1, S_2)$	21.426	22.436
$1, S_i, S_i^2, S_i^3, S_1 S_2, S_1^2 S_2, S_1 S_2^2$	21.223	22.299

Possiamo notare dalle tabelle che l'utilizzo di basi diverse porta un prezzo molto simile in tutti e tre i casi considerati.

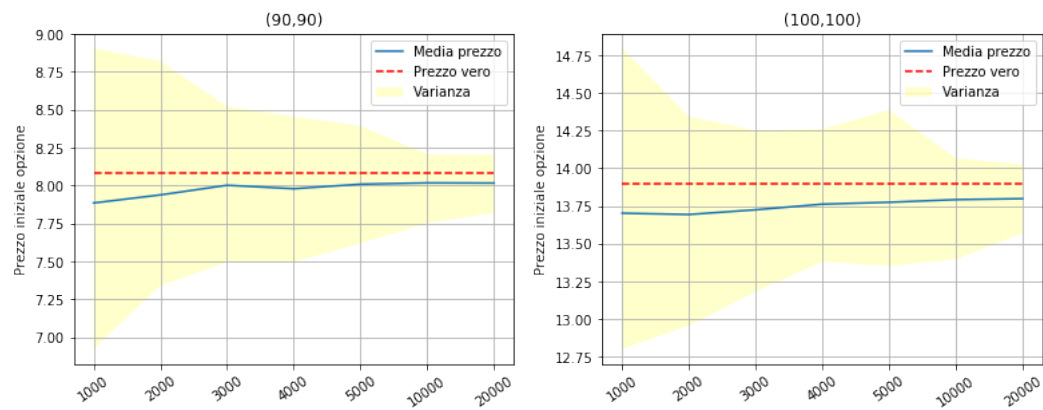
Ora utilizzando $1, S_i, S_i^2, S_i^3, h(S_1, S_2), i = 1, 2$ come basi per la regressione sono state fatte 100 repliche dell'algoritmo al crescere di b . Di seguito vengono riportati i grafici con la media del prezzo stimato per ognuno dei tre valori iniziali dei sottostanti al variare di b . La linea tratteggiata rappresenta il prezzo stimato utilizzando gli alberi binomiali. Per i "prezzi veri" si veda [5]. Per ogni valore di b sono state calcolati il massimo e il minimo valore stimato sulle 100 repliche. Dunque l'area gialla rappresenta l'intervallo dove sono contenuti tutti i valori stimati del prezzo iniziale che derivano dalle 100 repliche. Nel generare le traiettorie è stato utilizzato il metodo delle variabili antitetiche per ridurre la varianza.

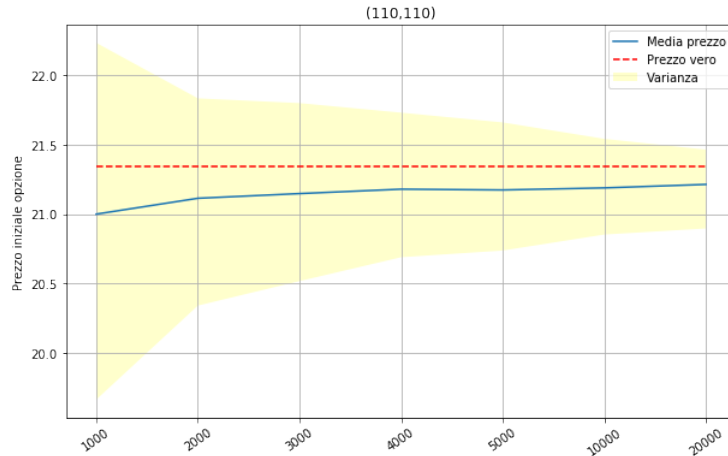
Gli esperimenti sono stati fatti utilizzando sia il LSM sia il LOOLSM.

LSM



LOOLSM

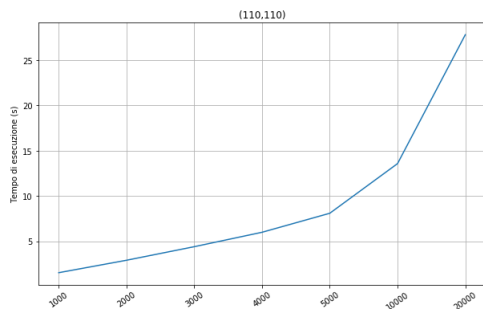




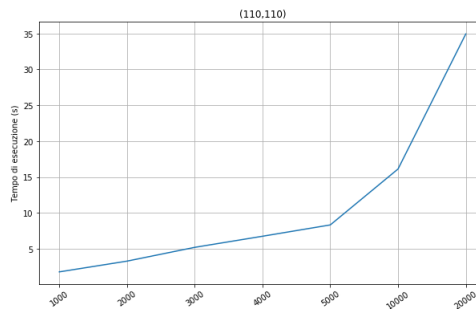
Dai grafici possiamo notare che per quanto riguarda il LSM l'high bias risulta essere ancora trascurabile al crescere di b . Nel caso del LOOLSM vediamo invece come ci aspettavamo che l'high bias è stato eliminato, infatti abbiamo un valore della media del prezzo stimato minore del prezzo vero, cioè rimane solo il bias subottimale.

È stato fatto anche un confronto dal punto di vista del costo computazionale tra i due metodi. Di seguito è riportato il tempo di esecuzione delle 100 replicazioni per ogni valore di b per entrambi i metodi nel caso in cui $S_1(0) = S_2(0) = 110$.

LSM



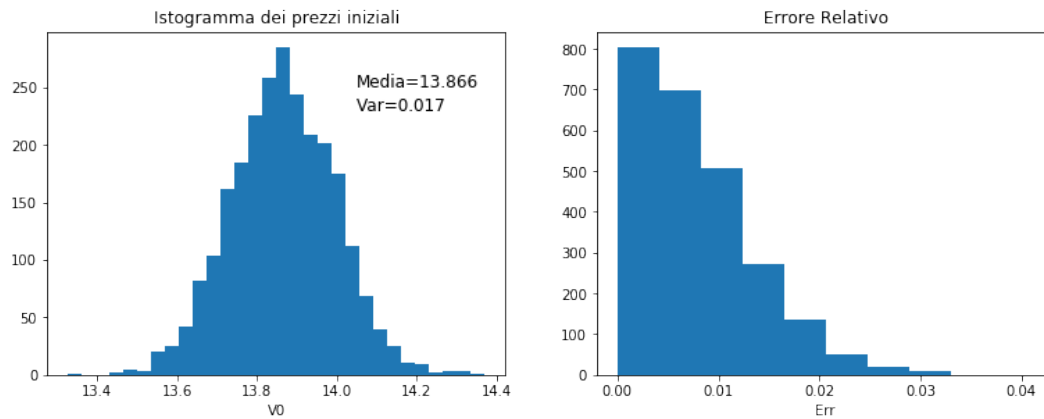
LOOLSM



Notiamo che il costo computazionale per un numero basso di simulazioni risulta pressochè uguale. Al crescere del numero di simulazioni LOOLSM come ci aspettiamo risulta avere un costo maggiore.

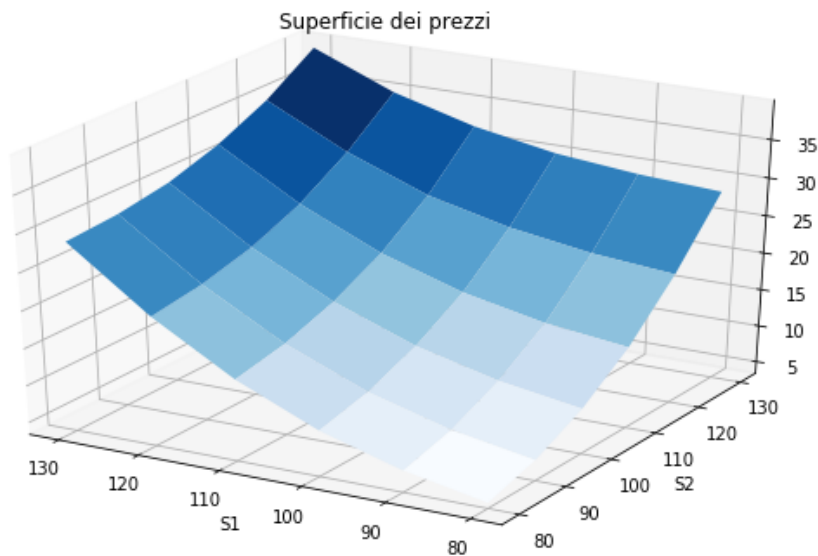
Ora fissando $S_1(0) = S_2(0) = 100$, $b = 10000$ e utilizzando sempre le basi precedenti vogliamo capire come sono distribuite le stime del prezzo iniziale dell'opzione che derivano dall'algoritmo LSM. Per fare ciò sono state operate 2500 replicazioni dell'algoritmo.

Di seguito vengono presentati 2 istogrammi. Il primo indica la frequenza dei prezzi iniziali stimati dall'algoritmo mentre il secondo indica l'errore relativo fatto rispetto al "prezzo vero" calcolato con il modello binomiale.



Dal secondo istogramma possiamo notare che il 95% delle replicazioni ha un errore relativo minore di 0.025.

Di seguito c'è un grafico che presenta la superficie dei prezzi dell'opzione in funzione dei valori iniziali dei sottostanti. Ogni stima è stata fatta con la media su 100 replicazioni dell'algoritmo LSM con $b = 10000$.



2.2.2 Massimo di due sottostanti correlati

Consideriamo lo stesso payoff dell'esperimento precedente con i due sottostanti sono modellizzati con dei moti Browniani geometrici in questo caso correlati. Sono stati considerati i seguenti parametri:

$$K = 100 \quad r = 0.05 \quad \delta = 0.1$$

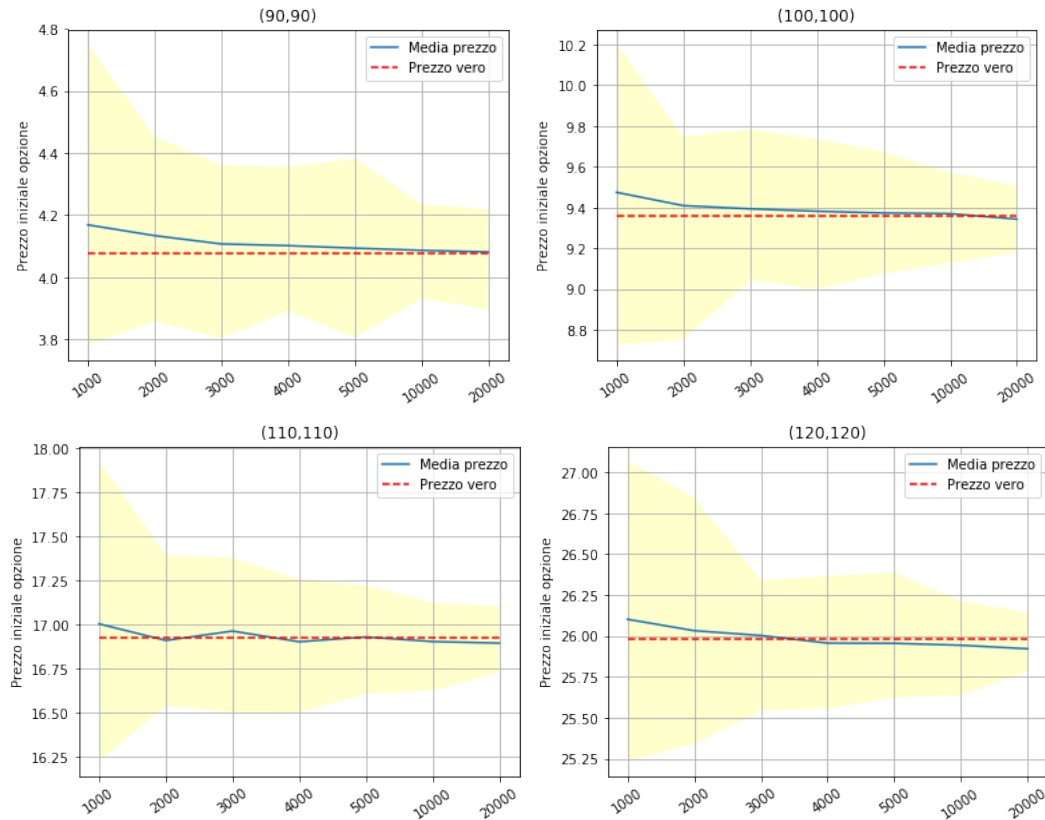
$$\sigma = 0.2 \quad \rho = 0.3 \quad b = 10000$$

L'opzione scade a $T = 1$ e può essere esercitata in nei tempi $t_i = i/3$ con $i = 1, 2, 3$. Per la regressione sono state usate le solite basi. Di seguito è riportata una tabella con il prezzo stimato attraverso il LSM utilizzando diversi valori iniziali dei sottostati.

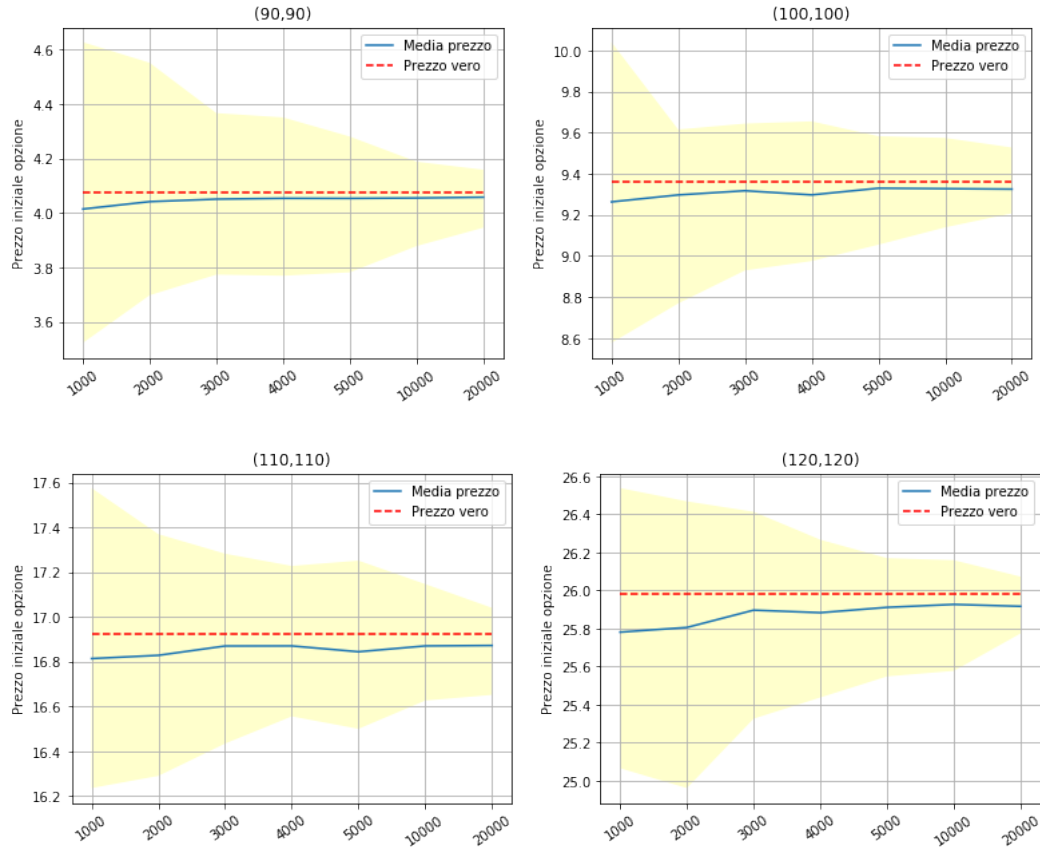
S_0^1	S_0^2	LSM
100	100	9.315
90	90	4.052
100	90	7.010
110	90	12.047

Come nel caso dei sottostanti non correlati è stato fatto un grafico che rappresenta la media della stima dei prezzi fatta su 100 replicazioni. La linea tratteggiata indica il prezzo calcolato mediante l'utilizzo di modelli multivariati. Per i prezzi si fa riferimento a [2].

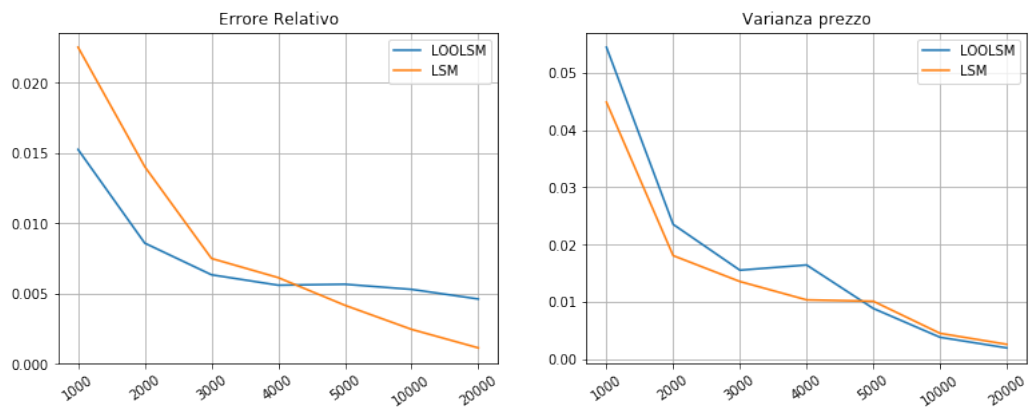
LSM



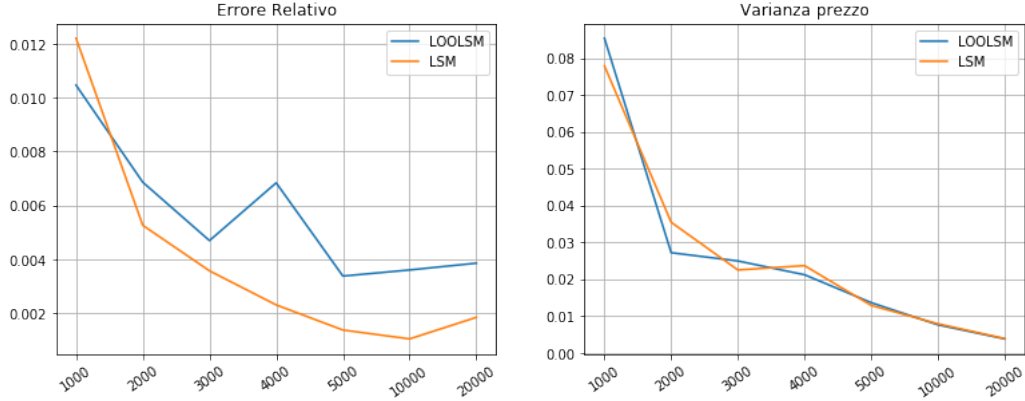
LOOLSM



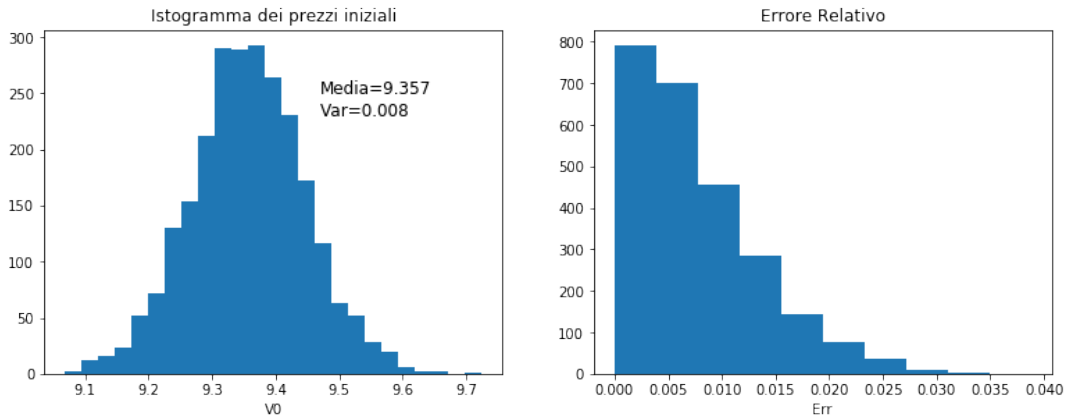
Anche qui si vede la presenza dell'high bias per il LSM che risulta trascurabile al crescere di b . Vediamo ora l'errore relativo della media dei valori iniziali stimati sulle 100 replicazioni rispetto al prezzo calcolato con i modelli multivariati e la varianza nei casi $S_i(0) = 90, 100$.



$$S_1(0) = S_2(0) = 100$$



Come per il caso indipendente per $S_1(0) = S_2(0) = 100$ vengono riportati in un istogramma i prezzi stimati di 2500 repliche dell'algoritmo LSM con $b = 20000$ e l'errore relativo di ogni replicazione del prezzo stimato rispetto al prezzo "vero" uguale a 9.361.



Notiamo che tutte più del 95% delle repliche hanno un errore relativo minore di 0.02.

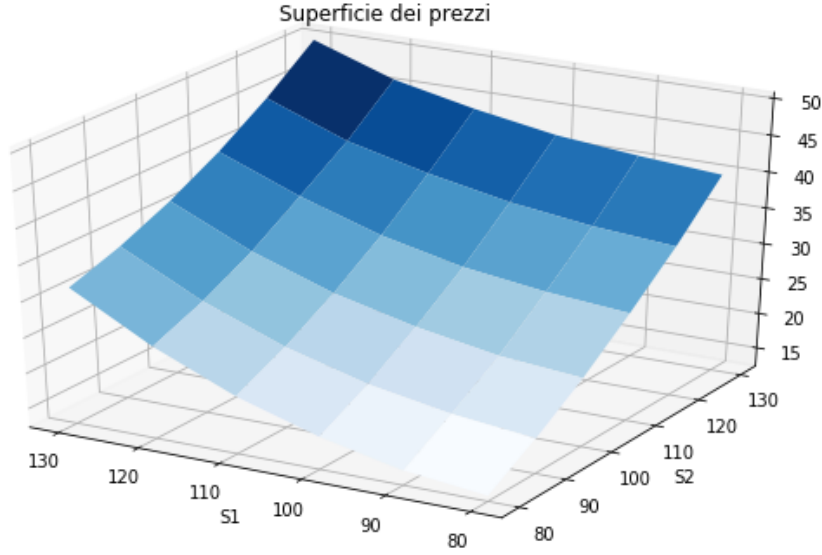
Vediamo ora la superficie della stima dei prezzi iniziali fatta con la media su 100 repliche dell'algoritmo LSM con $b = 10000$. In questo caso sono stati considerati i seguenti parametri:

$$T = 3 \quad m = 9 \quad K = 100 \quad r = 0.05$$

Supponiamo che i due sottostanti modellizzati con dei moti Browniani geometrici con volatilità e tassi di dividendo diversi:

$$\sigma_1 = 0.2 \quad \sigma_2 = 0.4$$

$$\delta_1 = 0.1 \quad \delta_2 = 0.08 \quad \rho = 0.3$$



2.2.3 Opzione Call su Basket di quattro sottostanti

Consideriamo un'opzione bermudiana su quattro sottostanti con la funzione payoff:

$$h(S_1, S_2, S_3, S_4) = \left(\frac{S_1 + S_2 + S_3 + S_4}{4} - K \right)^+$$

I sottostanti sono modellizzati con dei moti Browniani correlati con i seguenti parametri:

$$S_j(0) = 100, \quad \sigma_j = 0.4, \quad r = \delta_j = 0, \quad \rho = 0.5 \quad j = 1, 2, 3, 4$$

$$T = 5, \quad m = 10, \quad t_i = \frac{i}{2} \quad i = 1, 2, \dots, 10$$

Sono stati fatti esperimenti numerici con diversi valori di $K = 60, 80, 100, 120$. Siccome i sottostanti non pagano i dividendi, allora è ottimale non esercitare l'opzione fino alla scadenza. Dunque il prezzo dell'opzione bermudiana è uguale al prezzo di un'opzione europea. Per gli esperimenti sono state usate come basi per la regressione:

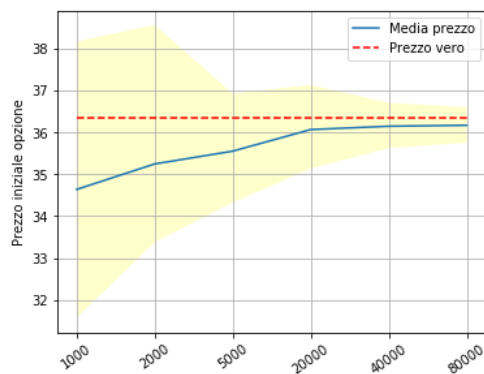
$$\psi(S) = (1, h(S), S_i \cdots, S_i^2, \dots, S_i S_j, \dots) \quad 1 \leq i < j \leq 4$$

Per un totale di 16 basi. Di seguito abbiamo una tabella che rappresenta la media della stima dei prezzi fatta su 100 repliche degli algoritmi, ognuna utilizzando 20000 simulazioni. È indicato anche il prezzo vero ottenuto da Choi [12].

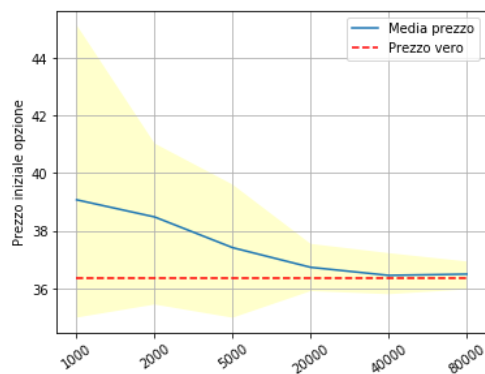
K	LSM	LOOLSM	Prezzo vero
60	47.927	47.196	47.481
80	36.762	36.067	36.352
100	28.481	27.750	28.007
120	22.104	21.402	21.763

Possiamo notare dalla tabella la presenza massiccia dell'high bias nell'LSM. Vediamo ora la convergenza della media della stima dei prezzi stimati nel caso in cui $K = 80$.

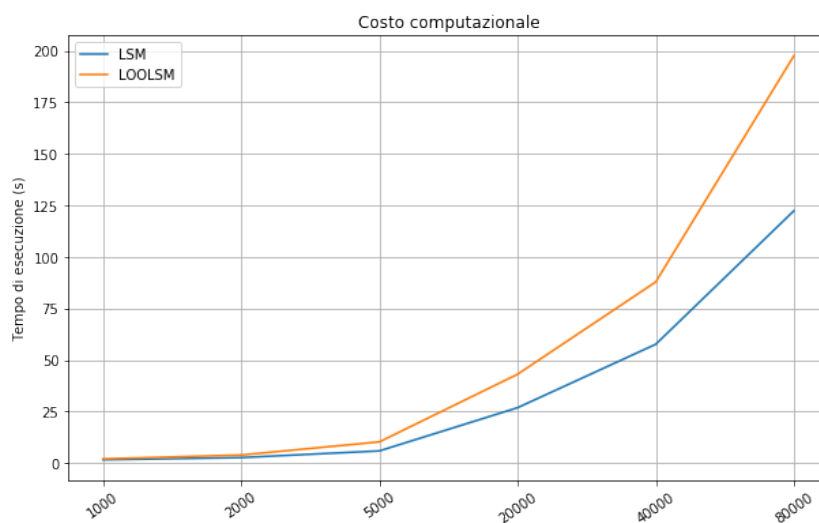
LOOLSM



LSM



Rispetto agli esperimenti con uno o due sottostanti si nota bene che nel caso LSM i prezzi stimati sono più distribuiti sopra il prezzo vero, mentre nel caso LOOLSM sono distribuiti più sotto. In entrambi i casi comunque per un numero alto di simulazioni la media si avvicina di molto al "prezzo vero". In questo caso è stato studiato il costo computazionale dei due algoritmi.



Per quanto riguarda il costo computazionale dell'LSM, esso è molto simile a quello visto per il caso dei 2 sottostanti. Mentre LOOLSM ha un costo che diventa molto elevato al crescere del numero di simulazioni. Risulta chiaro che se si vuole eliminare l'high bias e avere una stima accettabile del prezzo si deve prevedere un alto costo computazionale. Ricordiamo che LSM considera solo le traiettorie "in the money" a differenza dell'LOOLSM e inoltre quest'ultimo richiede il calcolo della diagonale di H , dunque il crescere del numero di basi e del numero di traiettorie richiede un tempo di esecuzione più elevato.

2.3 Convergenza coefficienti di regressione

Vediamo ora qualche esperimento sulla convergenza dei coefficienti di regressione dell'algoritmo LSM per il caso di un singolo periodo.

È stata replicata 5000 volte la stima dei coefficienti di regressione, per ogni replicazione è stato calcolato $|\beta - \hat{\beta}|^2$ e poi è stata fatta la media. Per i nostri test abbiamo scelto $\beta = (0, 0, \dots, 1)^T$ e $Y = e^{-N^2 \frac{t_2}{2} + t_2 W_{t_2}}$ e abbiamo posto $t_1 = 1/10$ e $t_2 = 1/5$. Di seguito viene riportata la tabella con le stime di $MSE(\hat{\beta})$ in base al numero di simulazioni b e al numero di basi N .

In fondo sono calcolati i due valori critici $V_{c1} = \sqrt{\frac{\log N}{5t_1 + t_2}}$ e $V_{c2} = \sqrt{\frac{\log N}{3t_1 + t_2}}$.

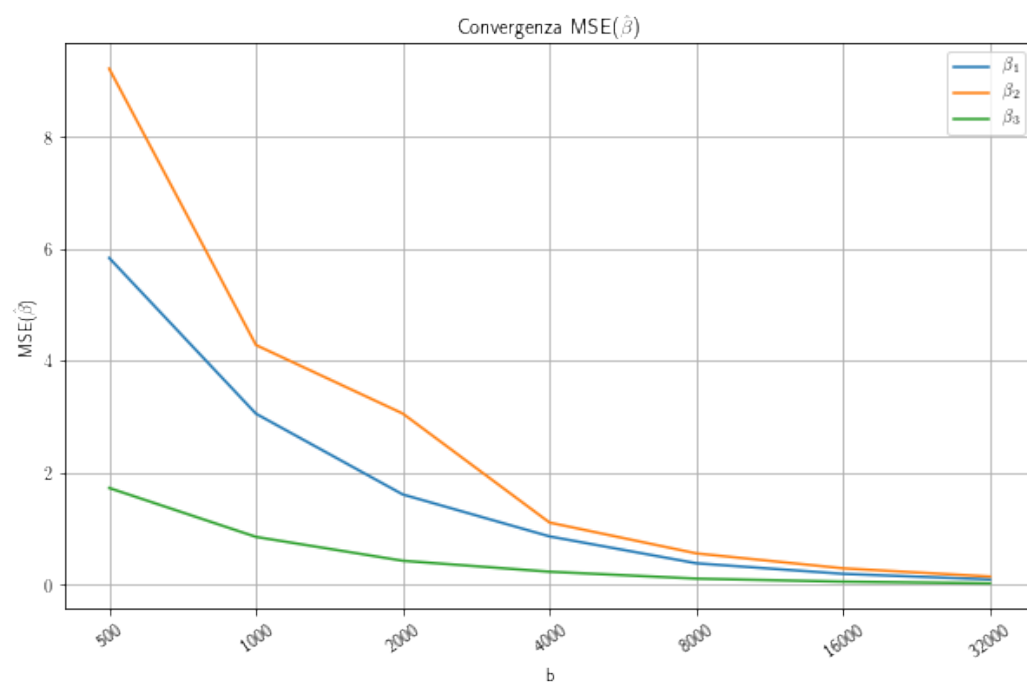
b								
N	500	1000	2000	4000	8000	16000	32000	64000
1	0.0788	0.0403	0.0203	0.0098	0.0052	0.0025	0.0013	0.0006
2	13.234	8.693	4.127	1.7223	0.9525	0.4646	0.2405	0.1243
3	4276.89	2605.58	1555.96	877.56	338.50	265.82	103.76	56.241
V_{c1}	2.98	3.14	3.29	3.44	3.58	3.72	3.85	3.98
V_{c2}	3.53	3.72	3.9	4.07	4.24	4.4	4.55	4.7

Dalla tabella notiamo che per $N = 3$ la convergenza è molto lenta e richiede un numero di simulazioni molto elevato. Seppure non riportati in tabella sono stati fatti esperimenti anche per $N > 3$ e la convergenza risulta sempre molto lenta.

Fissato $N = 2$ ho studiato la convergenza di $MSE(\hat{\beta})$ per tre differenti vettori β tale che $|\beta| = 1$. Abbiamo sempre utilizzato $t_1 = 1/10$ e $t_2 = 1/5$. I vettori β che abbiamo considerato sono i seguenti:

$$\beta_1 = \left[\frac{1}{3}, \frac{1}{12}, \frac{7}{12} \right]^T \quad \beta_2 = \left[\frac{1}{5}, 0, -\frac{4}{5} \right]^T \quad \beta_3 = \left[0, -\frac{1}{2}, \frac{1}{2} \right]^T$$

Di seguito sono riportati i tre grafici di convergenza di $MSE(\hat{\beta})$, uno per ogni β_i .



Appendice

L'attesa condizionata ha un ruolo estremamente importante in ambito finanziario. Vediamo alcune sue proprietà utili che sono state utilizzate anche in questo testo:

Per ogni $X, Y \in L^1(\Omega, \mathcal{F}, P)$, $a, b \in \mathbb{R}$ e $\mathcal{G} \subseteq \mathcal{F}$ si ha:

1. Se X è \mathcal{G} -misurabile $X = E[X|\mathcal{G}]$
2. Se X e \mathcal{G} sono indipendenti allora $E[X] = E[X|\mathcal{G}]$
3. $E[E[X|\mathcal{G}]] = E[X]$
4. $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$
5. Se $X \leq Y$ q.s. allora $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$
6. Se Y è indipendente da $\sigma(X, \mathcal{G})$ allora $E[XY|\mathcal{G}] = E[X|\mathcal{G}]E[Y]$
7. Se Y è \mathcal{G} -misurabile e limitata allora $E[XY|\mathcal{G}] = YE[X|\mathcal{G}]$
8. Se $\mathcal{H} \subseteq \mathcal{G}$ allora $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$
9. Sia φ una funzione convessa tale che $\varphi(X) \in L^1(\Omega, P)$ allora

$$E[\varphi(X)|\mathcal{G}] \geq \varphi(E[X|\mathcal{G}])$$

La sigma algebra di uno spazio di probabilità può essere visto come un'insieme di informazioni. In un processo stocastico siamo interessati al flusso di informazioni che nel tempo crescono. Per questo motivo abbiamo bisogno di introdurre la seguente:

Definizione. Una filtrazione $(\mathcal{F}_t)_{t \geq 0}$ su uno spazio di probabilità (Ω, \mathcal{F}, P) è una famiglia crescente di sotto- σ -algebre di \mathcal{F} .

Tra i processi stocastici continui sicuramente particolare attenzione si deve dare ai Moti Browniani che hanno diversi utilizzi in ambito stocastico. Vediamo la definizione e le proprietà che li caratterizzano.

Definizione. Sia $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ uno spazio di probabilità con filtrazione. Un moto Browniano d -dimensionale è un processo stocastico $W = (W_t)_{t \geq 0}$ in \mathbb{R}^d tale che:

- $W_0 = 0$ P -q.s.
- W è un processo continuo e adattato a \mathcal{F}_t
- per $0 \leq s < t$ si ha che la variabile aleatoria:

$$W_t - W_s \sim \mathcal{N}_{0, (t-s)I_d}$$

ed è indipendente da \mathcal{F}_s

Una classe molto importante di processi stocastici utili in ambito finanziario sono sicuramente le martingale. Di seguito diamo la seguente:

Definizione. Sia M un processo stocastico su $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ uno spazio di probabilità con filtrazione \mathcal{F}_t . Diciamo che M è una martingala se:

- $M \in L^1(\Omega, P)$ ed è adattato a \mathcal{F}_t
- Soddisfa:

$$M_s = E[M_t | \mathcal{F}_s] \quad \forall 0 \leq s \leq t$$

Notiamo che una martingala M è costante in media, infatti:

$$E[M_t] = E[E[M_t | \mathcal{F}_0]] = E[M_0] \quad \forall t \geq 0$$

Quando si decide di esercitare un'opzione americana il processo viene stoppato. Il tempo corrispondente a quando si vuole stoppare il processo è aleatorio, ma voglio che la scelta possa essere fatta in base alle informazioni disponibili fino a quel momento. Per questo viene introdotto il concetto di tempo d'arresto.

Definizione. Sia $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ uno spazio di probabilità con filtrazione. Una variabile aleatoria

$$\tau : \Omega \rightarrow [0, +\infty]$$

è un tempo d'arresto se

$$\{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \geq 0$$

Adesso vediamo come si può calcolare la forma differenziale di un processo stocastico che è funzione del tempo e di un moto Browniano.

Teorema (Formula di Itô). Sia $f = f(t, x_1, \dots, x_d) \in C^{1,2}(\mathbb{R}^{d+1})$. Allora $f(t, W_t)$ è un processo di Ito e vale:

$$df(t, W_t) = \partial_t f(t, W_t)dt + \sum_{i=1}^d \partial_{x_i} f(t, W_t)dW_t^i + \frac{1}{2} \sum_{i=1}^d \partial_{x_i x_i} f(t, W_t)dt$$

Il metodo Monte Carlo è un metodo che ha una convergenza molto lenta. Di solito bisogna fare anche 10^7 simulazioni per avere una bassa varianza e dunque avere convergenza. Questo numero elevato di simulazioni crea dei problemi a livello computazionale e dunque bisogna usare delle tecniche per ridurre la varianza e quindi velocizzare la convergenza del metodo.

Tra le tecniche di riduzione della varianza noi abbiamo usato negli esperimenti numerici quella delle variabili antitetiche. Questa tecnica consiste nel creare una correlazione tra le osservazioni. La tecnica è molto semplice, al posto di fare b simulazioni indipendenti l'idea è quella di operare solo $\frac{b}{2}$ simulazioni $y_1, \dots, y_{\frac{b}{2}}$ e prendere le restanti $\frac{b}{2}$ come $-y_1, \dots, -y_{\frac{b}{2}}$.

Se con Y indichiamo tutte le simulazioni e con Y^1 solo le prime $\frac{b}{2}$ si può dimostrare che:

$$Var(Y) \leq \frac{1}{2} Var(Y^1)$$

Bibliografia

- [1] Francis A. Longstaff and Eduardo S. Schwartz: Valuing american options by simulation : A simple least-square approach. *The Review of Financial Studies* Vol.14 No. 1, pp. 113-147 (2001).
- [2] M. Broadie and P. Glasserman: Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control* Volume 21, pp. 1323-1352 (1997).
- [3] E. Clement, D. Lamberton and P. Protter: An analysis of a least squares regression algorithm for American option pricing. *Finance and Stochastics*, pp. 449-471 (2002).
- [4] J. Tsitsiklis and B. Van Roy: Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, pp. 694-703 (2001).
- [5] P. Glasserman: Monte Carlo Methods in Financial Engineering. Cap. 8 pp. 421-479 (2003).
- [6] M. Moreno and Javier F. Navas: On the Robustness of Least - Squares Monte Carlo (LSM) for Pricing American Derivatives. *Review of Derivatives Research*, Vol. 6, No. 2 (2003).
- [7] A. Pascucci: Dispense dei corsi di Probabilità e Statistica Matematica e Analisi Stocastica.
- [8] A. Pascucci: Calcolo stocastico per la finanza. Springer-Verlag, Milano (2008).

- [9] A. Pascucci: PDE and Martingale Methods in Option Pricing. Bocconi University Press- Bocconi and Springer Series Italia, Milano (2011).
- [10] P. Glasserman and B. Yu: Number of paths versus number of basic functions in american option pricing. *The Annals of Applied Probability*, Vol. 14, No. 4, pp. 2090-2119 (2004).
- [11] J. Woo, C. Liu and J. Choi: Leave-one-out least square Monte Carlo algorithm for pricing american options (2018).
- [12] J. Choi: Sum of all Black-Scholes-Merton models: An efficient pricing method for spread, basket and Asian options. *Journal of Futures Markets*, pp. 627-644 (2018).